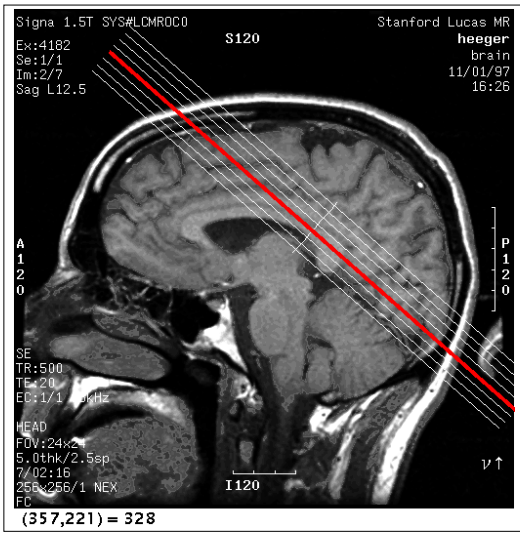
Mathematical Tools
for Neural and Cognitive Science

Fall semester, 2026

Section 6:

Decision-making & Categorization



Tumor, or not?

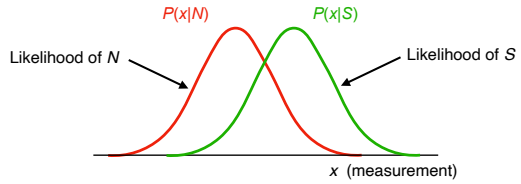
Decision-making and categorization (outline)

One-dimensional evidence, binary decision:
- Signal detection theory (SDT)
- Discriminability: Fisher Information

N-dimensional evidence, binary decision:
- Linear discriminant analysis (LDA)
- Quadratic discriminant analysis (QDA)

N-dimensional evidence, more than 2 categories:
- Labeled data: ML or MAP extension of QDA
- Unlabeled data: K-means or soft K-means clustering

Signal Detection Theory (or, Statistical Decision Theory)

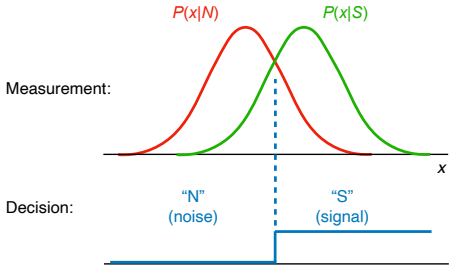


Stimulus is either “signal” (S) or “noise” (N).

P(*x* | *S*) and *P*(*x* | *N*) specify distributions of possible measurement x, conditioned on stimulus value.

After x is measured, an ideal observer uses these as “likelihood functions” of the stimuli value (*S* or *N*).

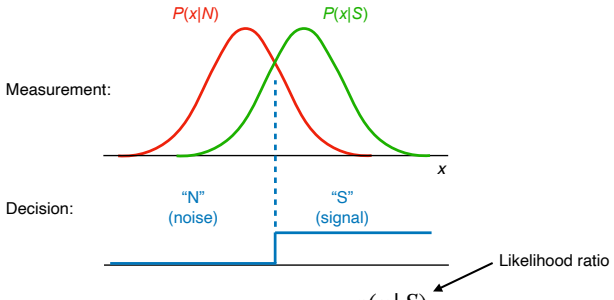
The Maximum likelihood (ML) decision rule



Say “S” if $p(x \mid S) > p(x \mid N)$

“N” otherwise.

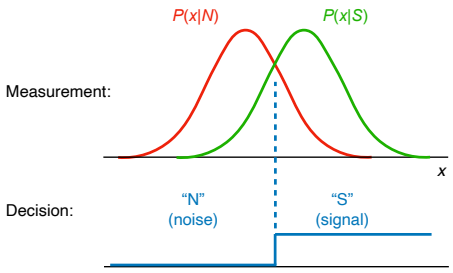
The Maximum likelihood (ML) decision rule



Say “S” if $p(x \mid S) > p(x \mid N)$, i.e., $\dfrac{p(x \mid S)}{p(x \mid N)} > 1$

“N” otherwise.

Maximum likelihood (ML) decision rule

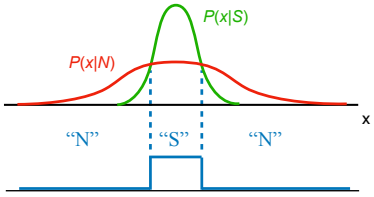


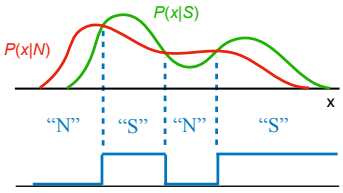
Say "S" if $x > \frac{\mu_S + \mu_N}{2} = c$

"N" otherwise.

(assuming equal-shaped symmetric unimodal distributions)

More generally, ML decision rule can have *multiple* thresholds:





Reminder: posterior via Bayes' Rule

Likelihood — Prior

Posterior

$$P(S \mid x) = \frac{p(x \mid S)P(S)}{p(x)}$$

normalizing term

The Maximum *a posteriori* (MAP) decision rule

Say "S" if $P(S|x) > P(N|x)$

"N" otherwise.

⇒ Say "S" if $\frac{p(x|S)P(S)}{p(x)} > \frac{p(x|N)P(N)}{p(x)}$

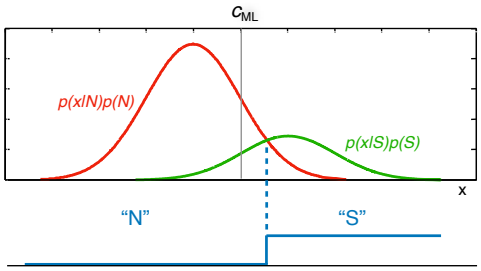
"N" otherwise.

⇒ Say "S" if $p(x|S)P(S) > p(x|N)P(N)$

"N" otherwise.

The MAP decision rule

maximizes proportion of correct answers, *taking prior probability into account*.



Compared to ML threshold, the MAP threshold moves *away* from higher-probability option.

Ratio form of MAP decision rule

Say "S" if $\frac{P(S|x)}{P(N|x)} > 1$

"N" otherwise, where

$$\frac{P(S|x)}{P(N|x)} = \left(\frac{p(x|S)}{p(x|N)}\right)\left(\frac{P(S)}{P(N)}\right)$$

"Posterior odds" "Likelihood ratio" "Prior odds"

Ratio form of MAP decision rule

Say "S" if $\frac{P(S|x)}{P(N|x)} > 1$

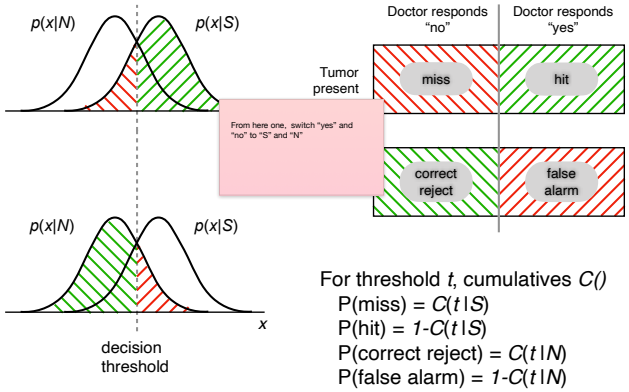
"N" otherwise, where

$$\frac{P(S|x)}{P(N|x)} = \left(\frac{p(x|S)}{p(x|N)}\right)\left(\frac{P(S)}{P(N)}\right)$$

Thus, say "S" if $\frac{P(x|S)}{P(x|N)} > \frac{P(N)}{P(S)}$

"Likelihood ratio" "Prior odds"

Signal Detection Theory: Potential outcomes



For threshold *t*, cumulatives *C()*

P(miss) = *C*(*t* | *S*)

P(hit) = *1-C*(*t* | *S*)

P(correct reject) = *C*(*t* | *N*)

P(false alarm) = *1-C*(*t* | *N*)

Bayesian decision rule

("maximum expected gain" or "minimum Bayes risk")

Incorporate *values* for the 4 possible outcomes:

"Payoff Matrix"

Stimulus	Response: No	Response: Yes
S	V_S^{No}	V_S^{Yes}
N	V_N^{No}	V_N^{Yes}

Bayes Optimal Criterion

$$\mathbb{E}(Yes\,|\,x) = V_S^{Yes}P(S\,|\,x) + V_N^{Yes}P(N\,|\,x)$$

$$\mathbb{E}(No\,|\,x) = V_S^{No}P(S\,|\,x) + V_N^{No}P(N\,|\,x)$$

Stimulus	Response: No	Response: Yes
S	V_S^{No}	V_S^{Yes}
N	V_N^{No}	V_N^{Yes}

Say "yes" if $\mathbb{E}(Yes\,|\,x) \geq \mathbb{E}(No\,|\,x)$

Say "yes" if $\dfrac{P(S\,|\,x)}{P(N\,|\,x)} \geq \dfrac{V_N^{No} - V_N^{Yes}}{V_S^{Yes} - V_S^{No}} = \dfrac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)}$

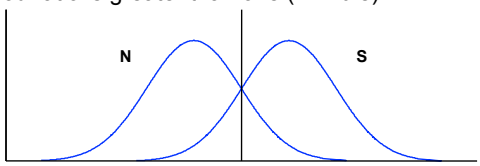
posterior odds

Bayes Optimal Criterion

Say "yes" if $\dfrac{P(S\,|\,x)}{P(N\,|\,x)} \geq \dfrac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)}$

i.e., if $\dfrac{p(x\,|\,S)}{p(x\,|\,N)} \geq \dfrac{P(N)}{P(S)}\dfrac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)} = \beta_{\text{opt}}$

Example, if equal priors and equal payoffs, say yes if the likelihood ratio is greater than one (ML rule):



Summary: Statistically optimal decision rules

(analogous to continuous estimation - see slides in previous section)

ML: Say "yes" if $\dfrac{p(x\,|\,S)}{p(x\,|\,N)} \geq 1$

MAP: Say "yes" if $\dfrac{p(x\,|\,S)}{p(x\,|\,N)} \geq \dfrac{P(N)}{P(S)}$

MEG: Say "yes" if $\dfrac{p(x\,|\,S)}{p(x\,|\,N)} \geq \dfrac{P(N)}{P(S)}\dfrac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)}$

The likelihood ratio is a "sufficient statistic".

Standardized SDT

Derivations of ML/MAP/MEG decision rules hold for *any* distributions, including different signal/noise distributions, discrete distributions (e.g., Poisson), and multi-dimensional distributions.

However, the standard SDT model that is most often used assumes equal-variance Gaussians in 1-D:

$$p(x|N) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_N)^2}{2\sigma^2}\right)$$

$$p(x|S) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_S)^2}{2\sigma^2}\right)$$

Standardized SDT

$$p(x|N) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_N)^2}{2\sigma^2}\right) \text{ and } p(x|S) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_S)^2}{2\sigma^2}\right)$$

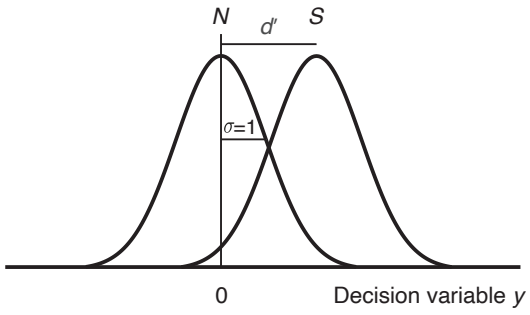
Let $y = \dfrac{x-\mu_N}{\sigma}$,

$$d' = \frac{\text{separation}}{\text{width}} = \frac{\mu_S - \mu_N}{\sigma}$$

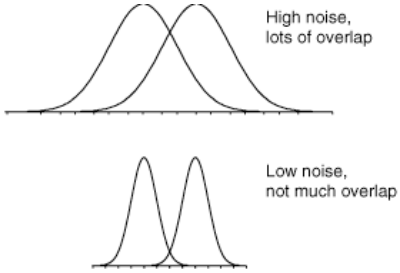
Then:

$$p(y|N) = \frac{1}{\sqrt{2\pi}} \exp\frac{-y^2}{2} \text{ and } p(y|S) = \frac{1}{\sqrt{2\pi}} \exp\frac{-(y-d')^2}{2}$$

Standardized SDT



Discriminability (d')

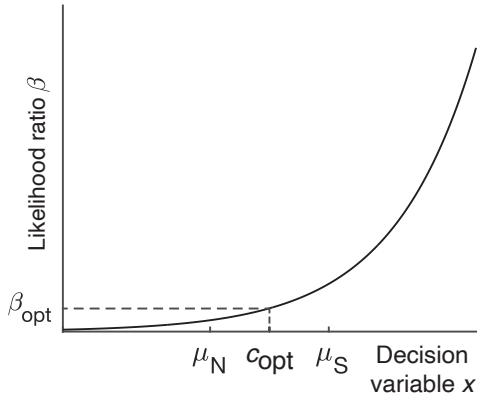


Standardized SDT

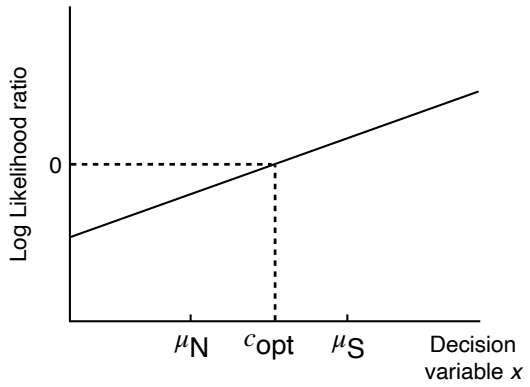
Likelihood ratio for $y = c$ is:

$$\frac{p(c\,|\,S)}{p(c\,|\,N)} = \frac{\exp\left[-\frac{(c-d')^2}{2}\right]}{\exp\left[-\frac{c^2}{2}\right]} = \exp\left[cd' - \frac{d'^2}{2}\right]$$

Standardized SDT



Standardized SDT



Standardized SDT

Likelihood ratio for $y = c$ is:

$$\beta_{opt} = \frac{p(c_{opt}\,|\,S)}{p(c_{opt}\,|\,N)} = \frac{\exp\left[-\frac{(c_{opt}-d')^2}{2}\right]}{\exp\left[-\frac{c_{opt}^2}{2}\right]} = \exp\left[c_{opt}d' - \frac{d'^2}{2}\right]$$

$$\log\beta_{opt} = c_{opt}d' - \frac{d'^2}{2}$$

$$c_{opt} = \frac{d'}{2} + \frac{\log\beta_{opt}}{d'}$$

$$= \frac{d'}{2} + \frac{1}{d'}\left[\log\frac{P(N)}{P(S)} + \log\frac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)}\right]$$

Standardized SDT

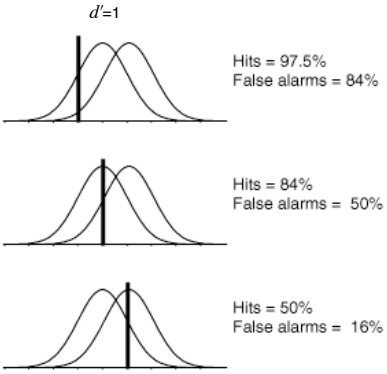
$$c_{opt} = \frac{d'}{2} + \frac{1}{d'}\left[\log\frac{P(N)}{P(S)} + \log\frac{V(\text{Correct}\,|\,N)}{V(\text{Correct}\,|\,S)}\right]$$

Optimal criterion is the ML criterion, shifted by a term that is a function of the prior odds plus a term that is a function of the payoff ratio.

Note: additivity of the effects of priors and payoffs is *not* seen in human behavior:

Locke, S. M., Gaffin-Cahn, E., Hosseinizaveh, N., Mamassian, P. & Landy, M. S. (2020). *Attention, Perception, & Psychophysics*, *82*, 3158-3175.

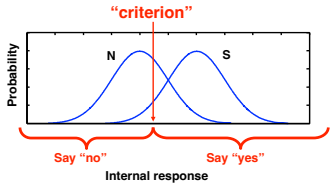
Signal Detection Theory: Criterion



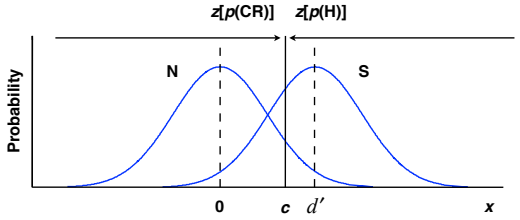
Example applications of SDT

- Vision
  - Detection (something vs. nothing)
  - Discrimination (lower vs greater level of: intensity, contrast, depth, slant, size, frequency, loudness, ...
- Memory (internal response = trace strength = familiarity)
- Neurometric function/discrimination by neurons (internal response = spike count)

From experimental measurements, assuming Gaussian distributions, can we determine the underlying values of d' and "criterion" (threshold)?



SDT: Estimating d' and c



$$d' = z[P(\text{Hit})] + z[P(\text{Correct Reject})]$$
$$= z[P(\text{Hit})] - z[P(\text{False Alarm})]$$
$$c = z[P(\text{Correct Reject})], \text{ where}$$
$$z(P) = \Phi^{-1}(P), \text{ where } \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp \frac{z^2}{2} dz$$

SDT: Estimating d' and c

$$d' = z[P(\text{Hit})] + z[P(\text{Correct Reject})]$$
$$= z[P(\text{Hit})] - z[P(\text{False Alarm})]$$
$$c = z[P(\text{Correct Reject})]$$

Stimulus	Response: No	Response: Yes
S	n_{Miss}	n_{Hit}
N	n_{CR}	n_{FA}

$$\hat{P}_{\text{Hit}} = n_{\text{Hit}}/(n_{\text{Hit}} + n_{\text{Miss}})$$
$$\hat{P}_{\text{FA}} = n_{\text{FA}}/(n_{\text{FA}} + n_{\text{CR}})$$

SDT: Estimating d' and c

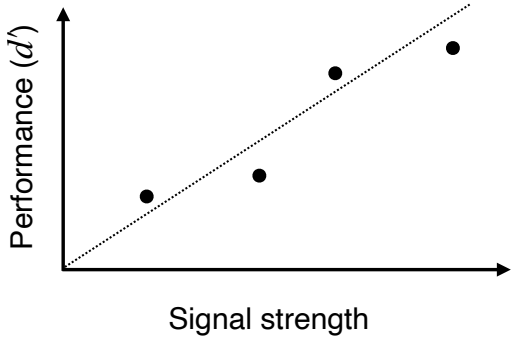
$$d' = z[P(\text{Hit})] + z[P(\text{Correct Reject})]$$
$$= z[P(\text{Hit})] - z[P(\text{False Alarm})]$$
$$c = z[P(\text{Correct Reject})]$$

Stimulus	Response: No	Response: Yes
S	$n_{\text{Miss}} + 0.5$	$n_{\text{Hit}} + 0.5$
N	$n_{\text{CR}} + 0.5$	$n_{\text{FA}} + 0.5$

$$\hat{P}_{\text{Hit}} = (n_{\text{Hit}} + 0.5)/(n_{\text{Hit}} + n_{\text{Miss}} + 1)$$
$$\hat{P}_{\text{FA}} = (n_{\text{FA}} + 0.5)/(n_{\text{FA}} + n_{\text{CR}} + 1)$$

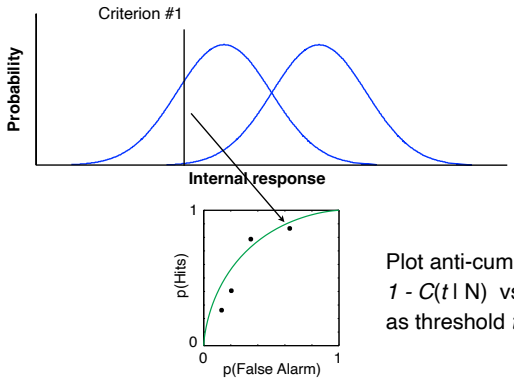
Hautus, M. J. (1995). Corrections for extreme proportions and the biasing effects on estimated values of d'. *Behavior Research Methods, Instruments, & Computers*, *27*, 46-51.

SDT: Psychometric function



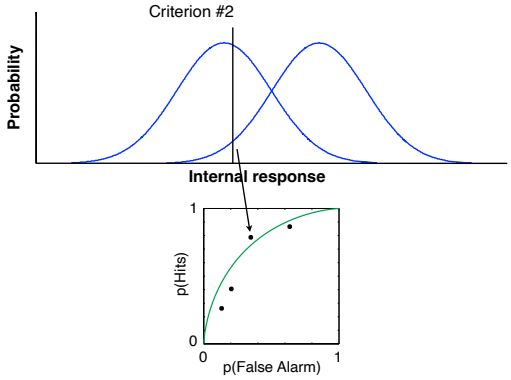
Note: 4 hit rates and one, shared, false-alarm rate

ROC (Receiver Operating Characteristic) curve

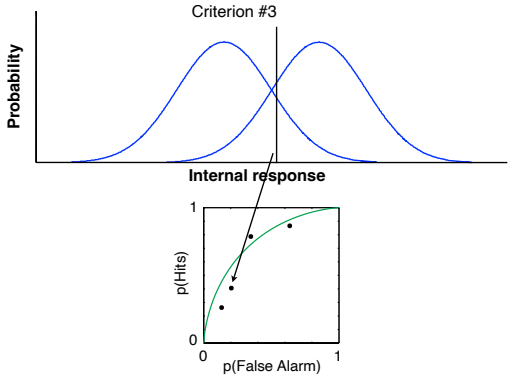


Plot anti-cumulatives:
1 - *C*(*t* | N) vs. *1*-*C*(*t* | *S*)
as threshold *t* varies

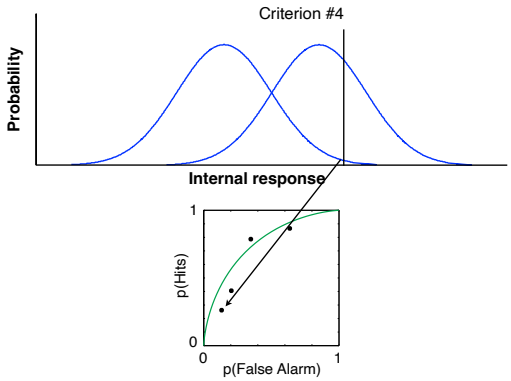
ROC (Receiver Operating Characteristic) curve



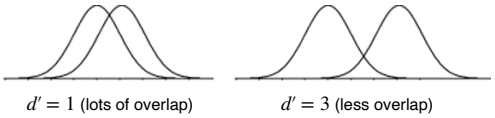
ROC (Receiver Operating Characteristic) curve



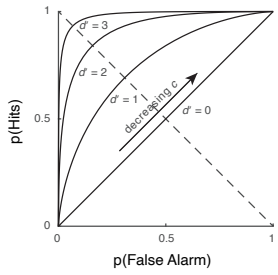
ROC (Receiver Operating Characteristic) curve



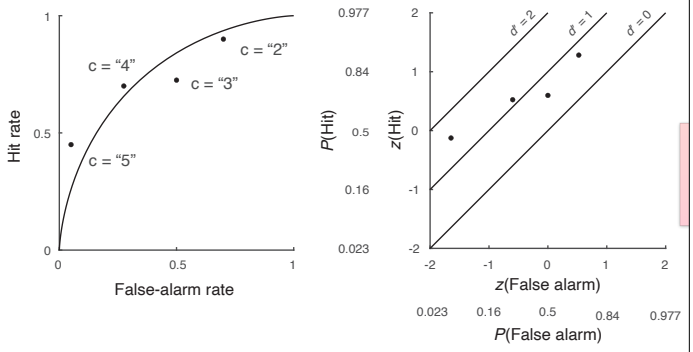
ROC (Receiver Operating Characteristic) curve



$d' = 1$ (lots of overlap) $d' = 3$ (less overlap)



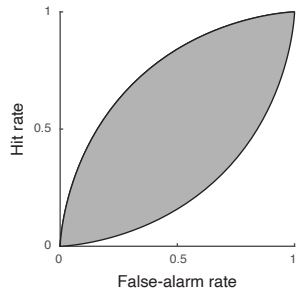
ROC (Receiver Operating Characteristic) curve



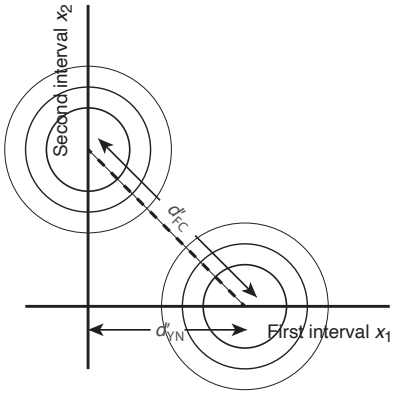
Or do a ML fit: Dorfman, D. D., & Alf, J., E. (1969). Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals: rating-method data. *Journal of Mathematical Psychology*, *6*, 487-496.

ROC (Receiver Operating Characteristic)

Achievable performance



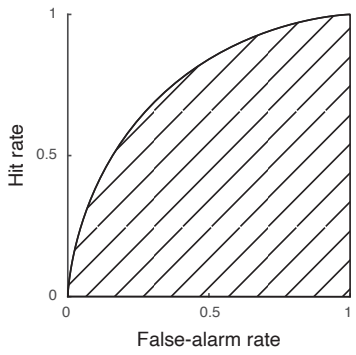
SDT and 2AFC



Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision Research*, *48*, 1837-1851.

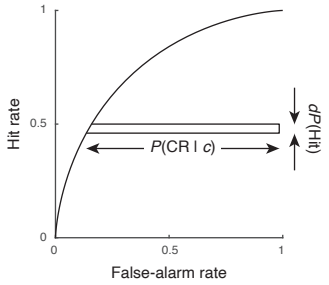
Area under the ROC curve

Area under curve = %correct in a 2AFC task!



Area under the ROC curve

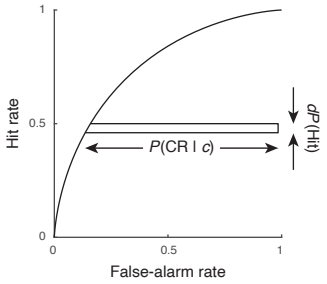
Area under curve = %correct in a 2AFC task

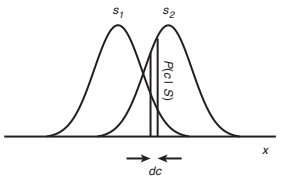


$$\text{AUROC} = \int_0^1 P(\text{Correct reject} \mid \text{criterion } c)\, dP(\text{Hit} \mid \text{criterion } c)$$

Area under the ROC curve

Area under curve = %correct in a 2AFC task





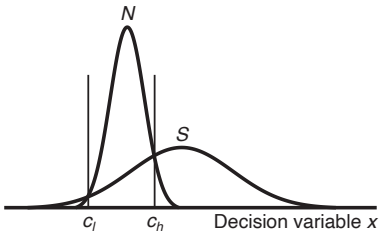
Slope of the ROC = likelihood ratio!

Area under the ROC curve

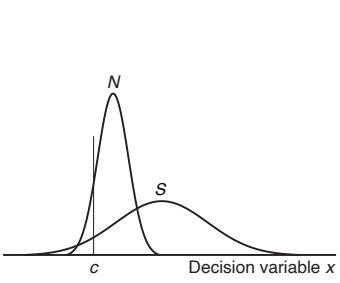
Area under curve = %correct in a 2AFC task

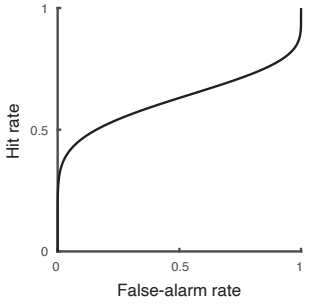
$$\begin{aligned}
\text{AUROC} &= \int_0^1 P(\text{Correct reject} \mid \text{criterion } c)\, dP(\text{Hit} \mid \text{criterion } c) \\
&= \int_{-\infty}^{\infty} p(x < c \mid N)\, p(\text{measurement is } c \mid S)\, dc \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{c} p(x \mid N)\, p(\text{measurement is } c \mid S)\, dx\, dc \\
&= \int_{-\infty}^{\infty} p(\text{measurement is } c \mid S) \int_{-\infty}^{c} p(x \mid N)\, dx\, dc \\
&= P_{\text{FC}}\,.
\end{aligned}$$

SDT with unequal variances

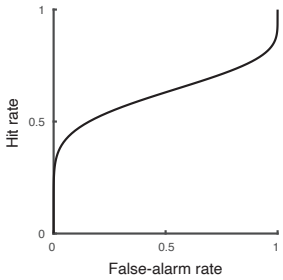


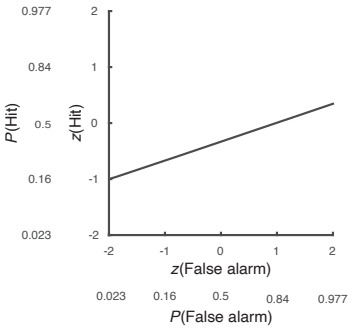
SDT with unequal variances



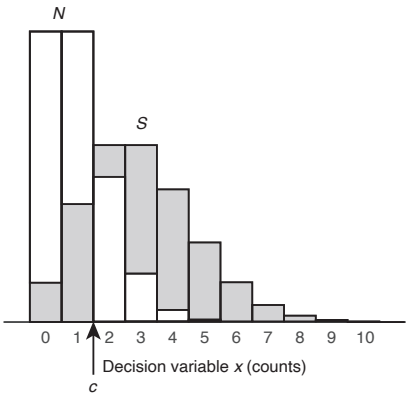


SDT with unequal variances

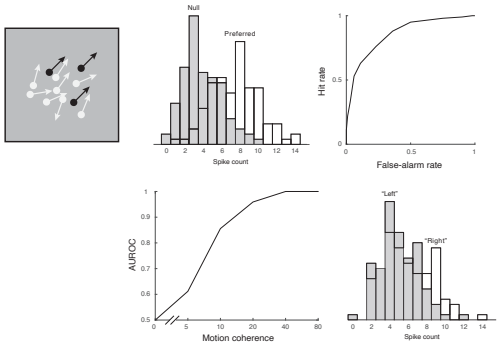




SDT with a discrete (Poisson) distribution



Area under the ROC - Poisson case or with data: Neurometric function and choice probability



Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, *12*, 4745-4765.
Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S. & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Visual Neuroscience*, *13*, 87-100.

Decision-making and categorization (outline)

One-dimensional evidence, binary decision:
- Signal detection theory (SDT)
- Discriminability: Fisher Information

N-dimensional evidence, binary decision:
- Linear discriminant analysis (LDA)
- Quadratic discriminant analysis (QDA)

N-dimensional evidence, more than 2 categories:
- Labeled data: ML or MAP extension of QDA
- Unlabeled data: K-means or soft K-means clustering

Fisher Information

- Second-order expansion of the (expected) negative log likelihood:

$$I(s) = -\mathbb{E}\left[\frac{\partial^2 \log p(r|s)}{\partial s^2}\right]$$

- Provides a bound on "precision" of unbiased estimators: (the "Cramér-Rao bound") $\sigma^2(s) \geq \frac{1}{I(s)}$

- Perceptually, provides a bound on **discriminability**: (Series et. al. 2009) $D(s) \leq \sqrt{I(s)}$

- Examples: with mean stimulus response $\mu(s)$

Gaussian case: $p(r|s) \sim \mathcal{N}(\mu(s), \sigma^2)$ $\quad I(s) = [\mu'(s)]^2/\sigma^2$

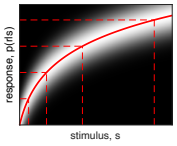
Poisson case: $p(r|s) \sim \text{Poiss}(\mu(s))$ $\quad I(s) = [\mu'(s)]^2/\mu(s)$

Example: Weber's law [Weber, 1834]

For many perceptual attributes, $D(s) \propto \frac{1}{s}$ (discrimination thresholds proportional to stimulus strength)

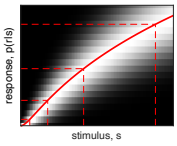
Assuming $I(s) \propto \frac{1}{s^2}$ what internal representation can explain this? Many!

additive Gaussian noise, with mean $\mu(s) = \log(s) + c$



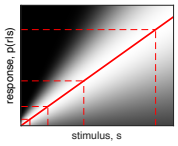
entirely due to response mean [Fechner, 1860]

Poisson noise, with mean $\mu(s) = [\log(s) + c]^2$



discrete representation, depends on both mean and variance

multiplicative Gaussian noise, with mean $\mu(s) = \alpha s$

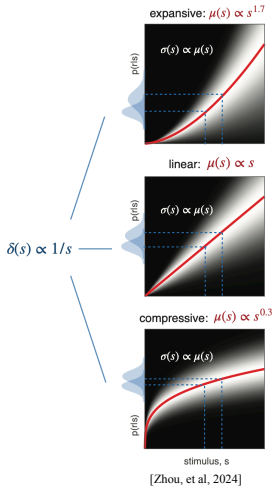


entirely due to response variance

[Zhou, et al, 2024]

S.S. Stevens. "To Honor Fechner and Repeal His Law: A power function, not a log function, describes the operating characteristic of a sensory system" (1961)

Three examples with different power-law mean response, each consistent with Weber's law discriminability.



[Zhou, et al, 2024]

Decision-making and categorization (outline)

One-dimensional evidence, binary decision:
- Signal detection theory (SDT)
- Discriminability: Fisher Information

N-dimensional evidence, binary decision:
- Linear discriminant analysis (LDA)
- Quadratic discriminant analysis (QDA)

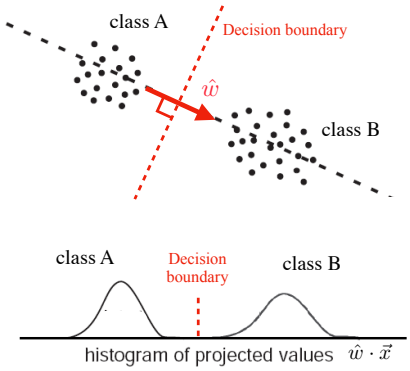
N-dimensional evidence, more than 2 categories:
- Labeled data: ML or MAP extension of QDA
- Unlabeled data: K-means or soft K-means clustering

Decision/classification in multiple dimensions

- Data-driven linear classifiers:
  - Prototype Classifier - minimize distance to class mean
  - Fisher Linear Discriminant (FLD) - maximize d'
  - Support Vector Machine (SVM) - maximize margin
- Statistical:
  - ML/MAP/Bayes under a probabilistic model
  - e.g.: Gaussian, identity covariance (same as Prototype)
  - e.g.: Gaussian, equal covariance (same as FLD)
  - e.g.: Gaussian, general case (Quadratic Discriminator)
- Some Examples:
  - Face classification
  - Neural population decoding

Linear Classifier

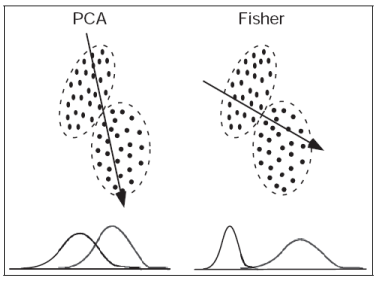
Find unit vector $\hat{w}$ ("discriminant") that best separates the distributions



Simplest linear discriminant: the Prototype Classifier

$$\hat{w} = \frac{\vec{\mu}_A - \vec{\mu}_B}{\|\vec{\mu}_A - \vec{\mu}_B\|}$$

Fisher Linear Discriminant



$$\max_{\hat{w}} \frac{\left[\hat{w}^T(\vec{u}_A - \vec{u}_B)\right]^2}{\left[\hat{w}^T C_A \hat{w} + \hat{w}^T C_B \hat{w}\right]}$$ (note: this is d-prime, squared!)

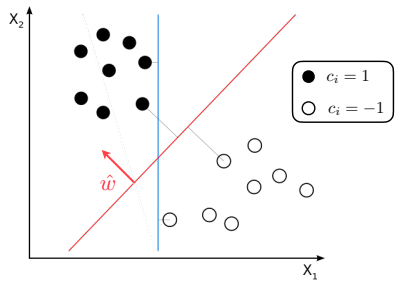
optimum: $\hat{w} = C^{-1}(\vec{u}_A - \vec{u}_B)$, where $C = \frac{1}{2}(C_A + C_B)$

Support Vector Machine (SVM)

(widely used in machine learning, but no closed form solution)

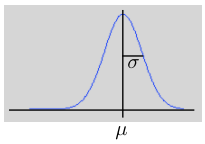
Maximize the "margin" (gap between data sets):

find largest m, and $\{\hat{w}, b\}$ s.t. $c_i(\hat{w}^T\vec{x}_i - b) \geq m, \quad \forall\ i$

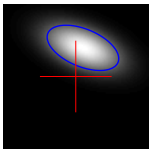


Reminder: Multi-D Gaussian densities

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



$$p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N |C|}}\, e^{-(\vec{x}-\vec{\mu})^T C^{-1} (\vec{x}-\vec{\mu})/2}$$

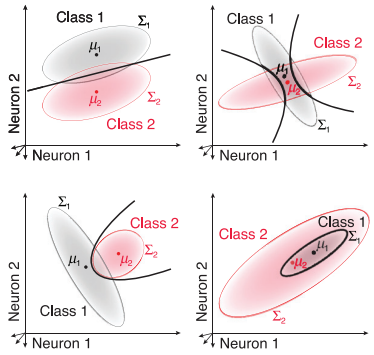


mean: [0.2, 0.8]
cov: [1.0 -0.3;
-0.3 0.4]

ML (or MAP) classifier for two Gaussians: QDA

Decision boundary is *quadratic*, with four possible geometries:

Simplest case: equal covariances



[on board]

[figure: Pagan et al. 2016]

A perceptual example: Biological gender identification (XX vs. XY)

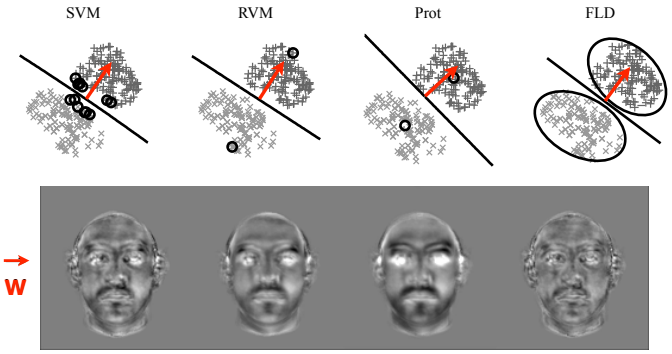




- 200 face images (100 male, 100 female)
- Adjusted for position, size, intensity/contrast
- Labeled by 27 human subjects

[Graf & Wichmann, NIPS*03]

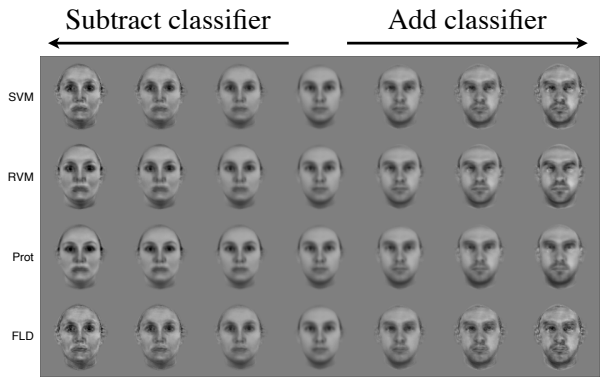
Linear classifiers



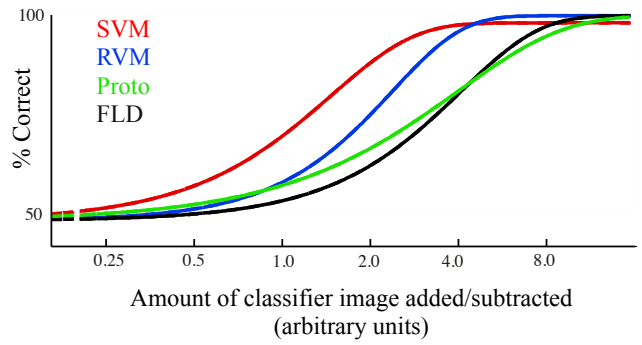
Four linear classifiers, trained on human data

Model validation/testing

- Cross-validation: Subject responses [% correct, reaction time, confidence] are explained
  - very well by SVM
  - moderately well by RVM / FLD
  - not so well by Prot
- Do these decision "models"make testable predictions? Synthesize optimally discriminable faces...



[Wichmann, et. al; NIPS*04]



[Wichmann, et. al; NIPS*04]

Decision-making and categorization (outline)

One-dimensional evidence, binary decision:
- Signal detection theory (SDT)
- Discriminability: Fisher Information

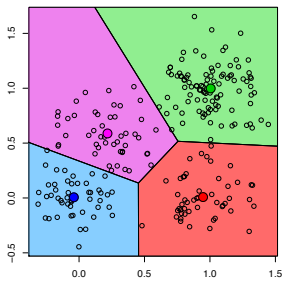
N-dimensional evidence, binary decision:
- Linear discriminant analysis (LDA)
- Quadratic discriminant analysis (QDA)

N-dimensional evidence, more than 2 categories:
- Labeled data: ML or MAP extension of QDA
- Unlabeled data: K-means or soft K-means clustering

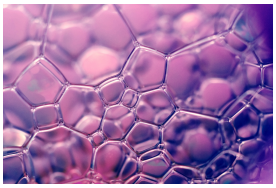
More than two categories, labeled data

If means and covariances are known, and the covariances are circular and identical across categories, this reduces to selecting the nearest neighbor:

Voronoi regions:

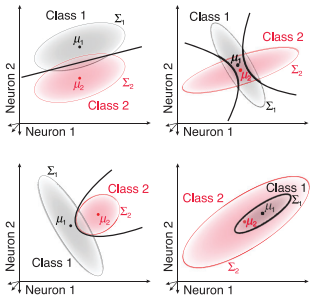


Soap bubbles:



Reminder: More than two categories, labeled data, Gaussian distributions (but not necessarily circular nor equal across categories).

ML (or MAP) classifier generalizes QDA:



[figure: Pagan et al. 2016]

Unlabeled data: Clustering

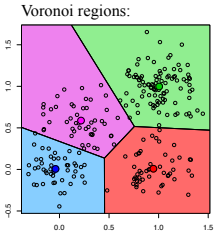
- K-Means (Lloyd, 1957)
- "Soft-assignment" version of K-means (a form of Expectation-Maximization - EM)
- In general, alternate between:
  1) Estimating cluster assignments (classification)
  2) Estimating cluster parameters
- Coordinate descent: converges to (possibly local) minimum
- Need to choose K (number of clusters) - cross-validation!

K-Means clustering algorithm

Alternate between two steps:

1. Estimate cluster assignments: given class centers, assign each point to closest one:

Voronoi regions:



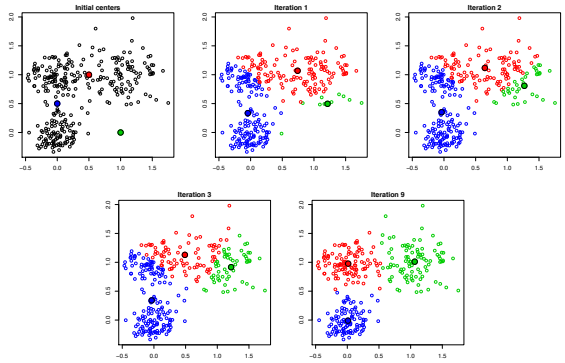
Soap bubbles:



2. Estimating cluster parameters: given assignments, re-estimate the centroid of each cluster.

K-means example

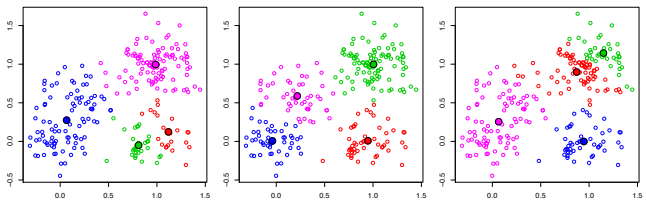
$N = 300$, and $K = 3$



[from R. Tibshirani, 2013]

K-means optimization failures

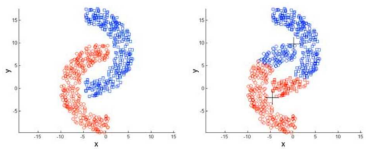
Initialization matters (due to local minima) …
Three solutions obtained with different random starting points:



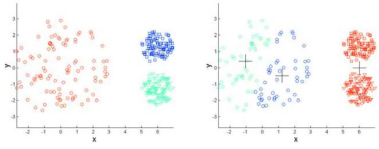
[from R. Tibshirani, 2013]

K-means systematic failures

Non-convex/non-round-shaped clusters



Clusters with different densities



Picture courtesy: Christof Monz (Queen Mary, Univ. of London)

ML for discrete mixture of Gaussians: soft K-means

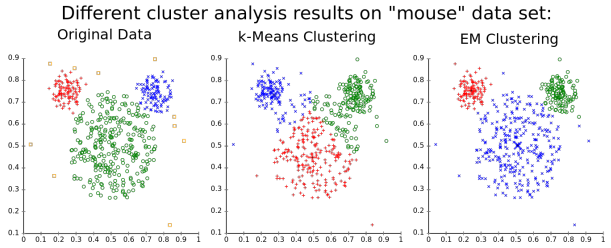
$$p(\vec{x}_n|a_{nk},\vec{\mu}_k,\Lambda_k) \propto \sum_k \frac{a_{nk}}{\sqrt{|\Lambda_k|}} e^{-(\vec{x}_n-\vec{\mu}_k)^T\Lambda_k^{-1}(\vec{x}_n-\vec{\mu}_k)/2}$$

a_{nk} = assignment *probability*

$\{\vec{\mu}_k, \Lambda_k\}$ = mean/covariance of class *k*

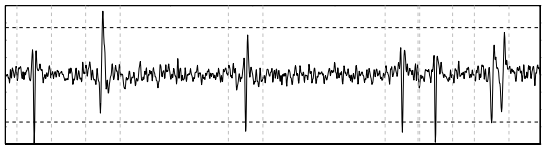
Intuition: alternate between maximizing these two sets of variables ("coordinate descent")

Essentially, a version of K-means with "soft" (i.e., continuous, as opposed to binary) assignments!



[wikipedia]

Application to neural "spike sorting"

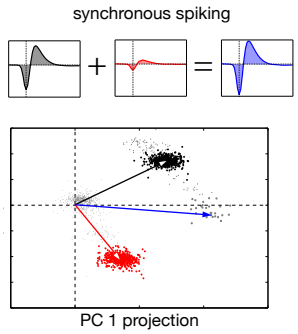


Standard solution:

1. Threshold to find segments containing spikes
2. Reduce dimensionality of segments using PCA
3. Identify spikes using clustering (e.g., K-means)

Note: Fails for overlapping spikes!

Failures of clustering for near-synchronous spikes

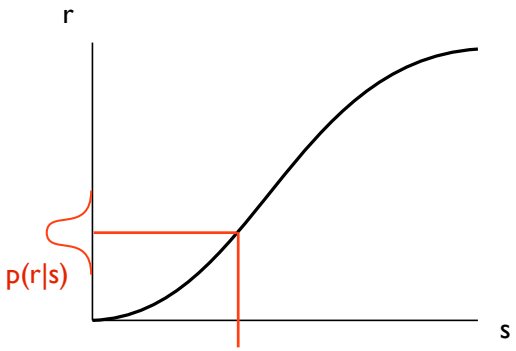


[Pillow et. al. 2013]

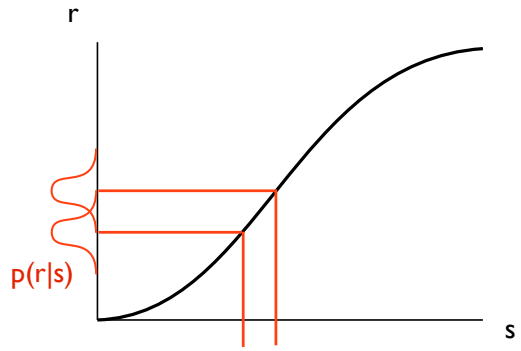
"Decoding" neural populations?

- Test/compare **encoding** models
- Connect neural response to behavior
- Build brain-computer interfaces

Encoding determines discriminability

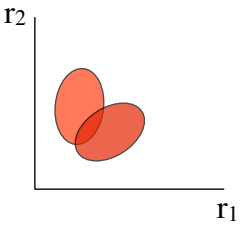


Probabilistic encoding model determines discriminability



Discriminability (d') is (approximately) slope/stdev

Two neurons



Same fundamental issues as 1D case:

- Probabilistic encoding determines discriminability
- Intuitively, overlap is distance/spread
- More precisely: estimate Fisher Information [on board]

I. Simple/intuitive population decoding

- Linear? $\hat{s}(\vec{r}) = \sum_n w_n r_n = \vec{w} \cdot \vec{r}$

  (simple, but usually doesn't work well)

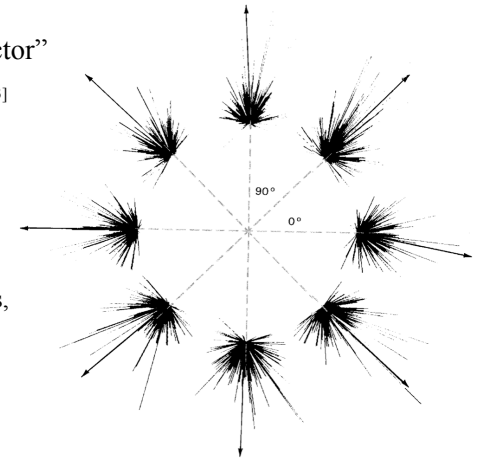
- Winner-take-all $\hat{s}(\vec{r}) = s_m, \qquad m = \arg\max_n \{r_n\}$

  (simple, but discontinuous and noise-susceptible)

- Population vector [Georgopoulos et.al., 1986] $\hat{s}(\vec{r}) = \frac{\sum_n r_n s_n}{\sum_n r_n}$

  (also simple, more robust)

"population vector"
[Kalaska, Caminiti Georgopoulous, 1983]



A sum of vectors, weighted by firing rate of motor neurons, predicts arm movement...

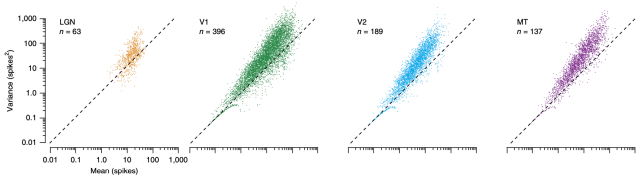
II. Statistically optimal decoding

- Maximum likelihood (ML) $\hat{s}(\vec{r}) = \arg\max_s p(\vec{r}|s)$
- Maximum a posteriori (MAP) $\hat{s}(\vec{r}) = \arg\max_s \; p(\vec{r}|s) \cdot p(s)$
- Minimum Mean Squared Error (MMSE), a.k.a. Bayes Least Squares (BLS) $\hat{s}(\vec{r}) = \mathbf{E}(s|\vec{r})$

Gaussian response noise?

Not a great model for neural noise, and ML estimation is nonlinear regression :(

Poisson response noise?



[Goris etal, 2014]

Better model for neural responses, and ML estimation is much nicer…

ML decoding for a Poisson-spiking neural population

[Ma, Beck, Latham, Pouget, 2006; Jazayeri & Movshon, 2006; Zhang et al, 1998]

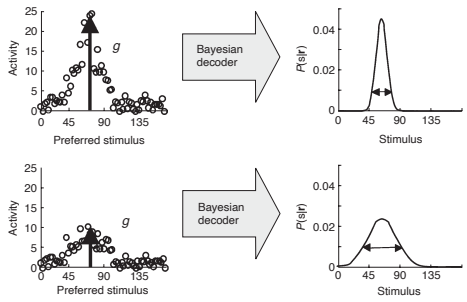
$$p(\vec{r}|s) = \prod_{n=1}^{N} \frac{h_n(s)^{r_n}\ e^{-h_n(s)}}{r_n!}$$

$$\log\left(p(\vec{r}|s)\right) = \sum_{n=1}^{N} r_n \log(h_n(s)) - h_n(s) - \log(r_n!)$$

If $\sum_{n=1}^{N} h_n(s)$ is constant (i.e., tuning curves "tile"), just minimize the response-weighted sum of log tuning curves.

Special cases allow closed-form solutions:

- Gaussian tuning curves $h_n(s) = \exp\left(-(s-s_n)^2/2\sigma^2\right)$
- von Mises tuning curves $h_n(s) = \exp\left(\kappa\cos(s-s_n)\right)$

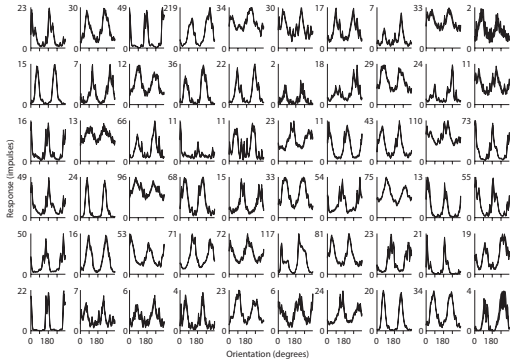


$$\hat{s}(\vec{r}) = \frac{\sum_n r_n s_n}{\sum_n r_n} \qquad \hat{\sigma}(\vec{r}) = \frac{\sigma_{TC}}{\sqrt{\sum r_n}}$$

[Ma, Beck, Latham & Pouget, 06]

Population decoding

The data: tuning curves f_i



[Graf, Kohn, Jazayeri & Movshon, 11]

Comparing population decoders

1) The ML decoder, assuming independent Poisson responses (the "PID"):

$$\log L(\theta) = \log\left(\prod_{i=1}^{N} p(r_i|\theta)\right) = \sum_{i=1}^{N} \log\left(\frac{f_i(\theta)^{r_i}}{r_i!}\exp(-f_i(\theta))\right)$$

$$= \sum_{i=1}^{N} \log(f_i(\theta))r_i - \sum_{i=1}^{N} f_i(\theta) - \sum_{i=1}^{N} \log(r_i!) = \sum_{i=1}^{N} W_i(\theta)r_i + B(\theta)$$

For discrimination between two values, likelihood ratio is a *linear* function of responses:

$$\log LR(\theta_1, \theta_2) = \log\left(\frac{L(\theta_1)}{L(\theta_2)}\right) = \log L(\theta_1) - \log L(\theta_2)$$

$$= \sum_{i=1}^{N} [W_i(\theta_1) - W_i(\theta_2)]r_i + [B(\theta_1) - B(\theta_2)]$$

$$= \sum_{i=1}^{N} w_i(\theta_1, \theta_2)r_i + b(\theta_1, \theta_2)$$

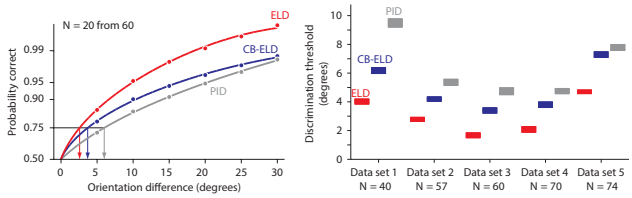
[Graf, Kohn, Jazayeri & Movshon, 11]

Comparing population decoders

2) Alternatively, compute an SVM on the measured response vectors for each orientation, the empirical linear decoder ("ELD"):

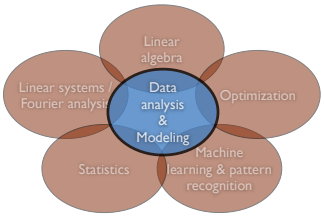
$$y(\theta_1, \theta_2) = \sum_{i=1}^{N} w_i(\theta_1, \theta_2)r_i + b(\theta_1, \theta_2)$$

3) For each neuron and orientation, shuffle the responses across trials and train a new SVM, the correlation-blind empirical linear decoder (CB-ELD).



[Graf, Kohn, Jazayeri & Movshon, 11]

Where we've been…



- Linear algebra / linear systems
  - Ex: Trichromacy
- Least squares
  - regression / TLS regression
- Linear shift-invariant systems
  - convolution / Fourier transforms
  - Ex: Auditory filtering
- Summary statistics - dispersion, central tendency, PCA
- Statistical inference & estimation
  - estimation: bias, variance, convergence
  - optimal estimation: ML, MAP, Bayes
  - model comparison, overfitting, regularization, cross-validation
  - Ex: fitting an LNP model
- Decision-making and categorization
  - Signal detection theory, Fisher information
  - classification, clustering
  - Ex: population decoding

Keep climbing!

