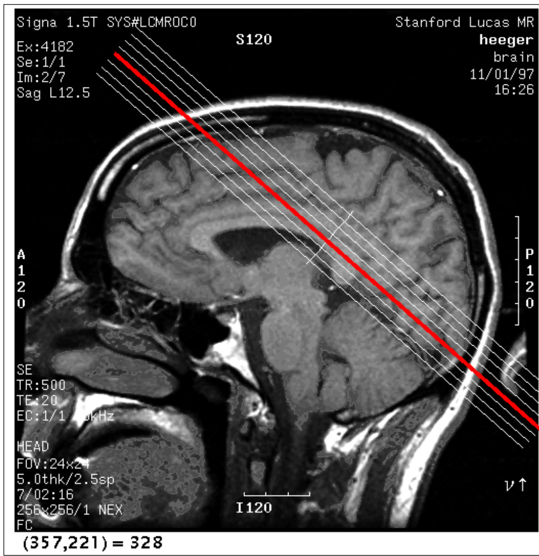
Mathematical Tools
for Neural and Cognitive Science

Fall semester, 2019

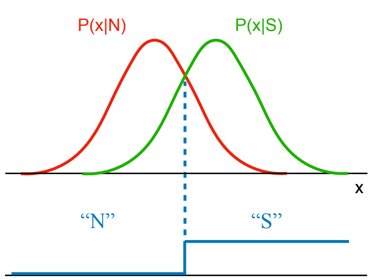
Section 5a:

Statistical Decision Theory
+
Signal Detection Theory



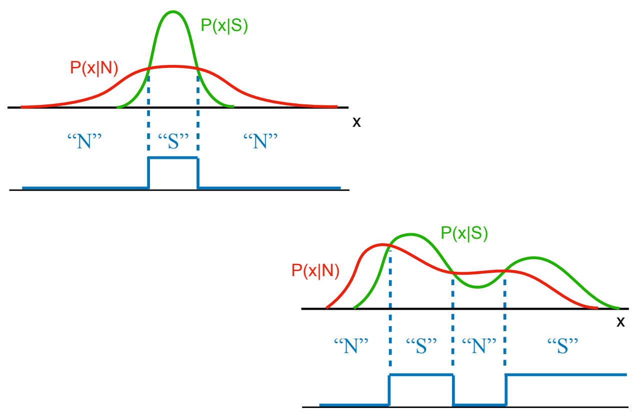

Tumor, or not?

## Signal Detection Theory (binary estimation)

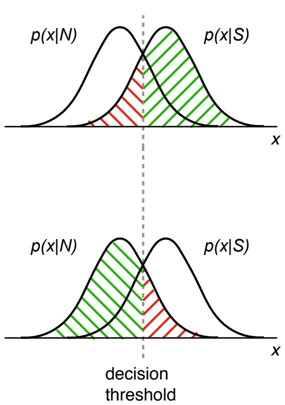


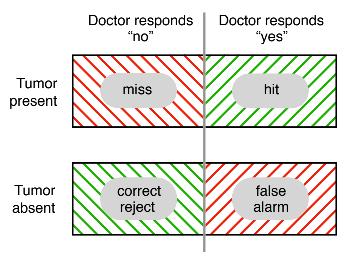

For equal-shape, unimodal, symmetric distributions,
the ML decision rule is a *threshold* function.

## More generally, decision rule can have multiple thresholds…




## Signal Detection Theory: Potential outcomes




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

*c(t|S)* = miss

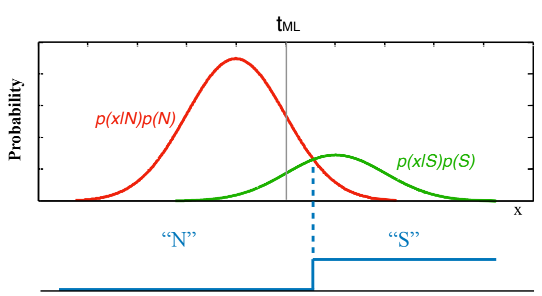
*1-c(t|S)* = hit

*c(t|N)* = correct reject

*1-c(t|N)* = false alarm

## MAP decision rule?

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




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

## Bayes decision rule?

Incorporate values for the four possible outcomes:

Payoff Matrix

| Stimulus | | Response: Yes | Response: No |
|---|---|---|---|
| | S+N | $V_{S+N}^{Yes}$ | $V_{S+N}^{No}$ |
| | N | $V_{N}^{Yes}$ | $V_{N}^{No}$ |

## Bayes Optimal Criterion

| Stimulus | | Response: Yes | Response: No |
|---|---|---|---|
| | S+N | $V_{S+N}^{Yes}$ | $V_{S+N}^{No}$ |
| | N | $V_{N}^{Yes}$ | $V_{N}^{No}$ |

$$E(Yes \mid x) = V_{S+N}^{Yes} p(S+N \mid x) + V_{N}^{Yes} p(N \mid x)$$
$$E(No \mid x) = V_{S+N}^{No} p(S+N \mid x) + V_{N}^{No} p(N \mid x)$$

Say yes if $E(Yes \mid x) \geq E(No \mid x)$

## Optimal Criterion

$$E(Yes \mid x) = V_{S+N}^{Yes} p(S+N \mid x) + V_{N}^{Yes} p(N \mid x)$$
$$E(No \mid x) = V_{S+N}^{No} p(S+N \mid x) + V_{N}^{No} p(N \mid x)$$

Say yes if $E(Yes \mid x) \geq E(No \mid x)$

Say yes if $\dfrac{p(S+N \mid x)}{p(N \mid x)} \geq \dfrac{V_{N}^{No} - V_{N}^{Yes}}{V_{S+N}^{Yes} - V_{S+N}^{No}} = \dfrac{V(\text{Correct} \mid N)}{V(\text{Correct} \mid S+N)}$

Posterior odds

# Apply Bayes' Rule

Posterior / Likelihood / Prior

$$p(S+N \mid x) = \frac{p(x \mid S+N)\,p(S+N)}{p(x)}$$

Nuisance normalizing term

$$p(N \mid x) = \frac{p(x \mid N)\,p(N)}{p(x)}, \text{ hence}$$

$$\frac{p(S+N \mid x)}{p(N \mid x)} = \left(\frac{p(x \mid S+N)}{p(x \mid N)}\right)\left(\frac{p(S+N)}{p(N)}\right)$$

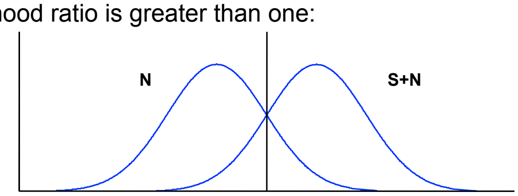
Posterior odds / Likelihood ratio / Prior odds

# Optimal Criterion

Say yes if $\dfrac{p(S+N \mid x)}{p(N \mid x)} \geq \dfrac{V(\text{Correct} \mid N)}{V(\text{Correct} \mid S+N)}$

i.e., if $\dfrac{p(x \mid S+N)}{p(x \mid N)} \geq \dfrac{p(N)}{p(S+N)} \dfrac{V(\text{Correct} \mid N)}{V(\text{Correct} \mid S+N)} = \beta$

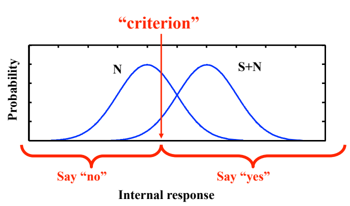
Example, if equal priors and equal payoffs, say yes if the likelihood ratio is greater than one:



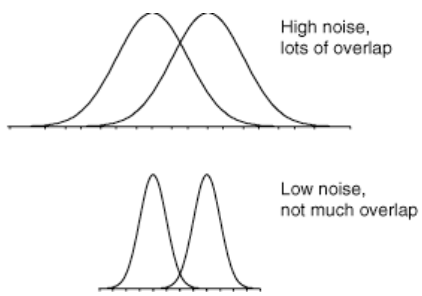

# Example applications of SDT

- 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)

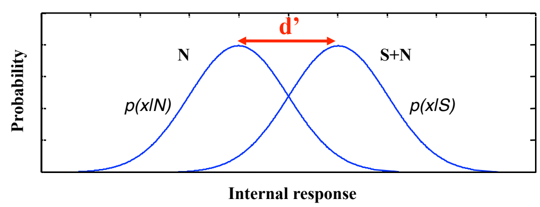
From experimental measurements, assuming human is optimal, can we determine the underlying distributions and criterion?



Signal Detection Theory: discriminability (d’)




Internal response: probability of occurrence curves

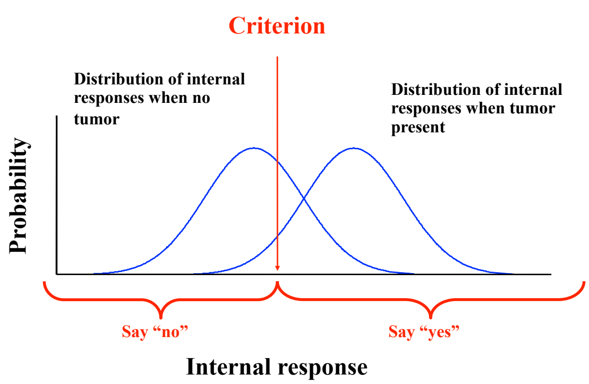



$$d' = \frac{\text{“separation”}}{\text{“width”}}$$

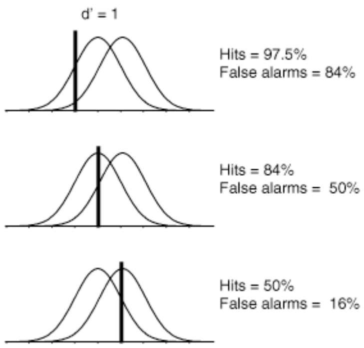
Discriminability (“d-prime”) is the normalized separation between the two distributions

Error rate is a function of d’

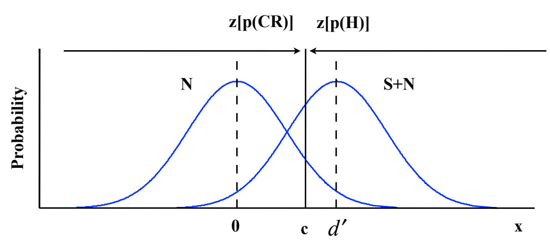
Criterion



## Signal Detection Theory: Criterion




## SDT: Gaussian case

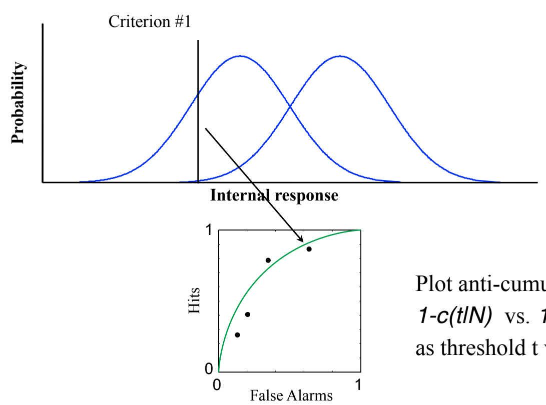



$$d' = z[p(H)] + z[p(CR)] = z[p(H)] - z[p(FA)]$$

$$c = z[p(CR)] \qquad G(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

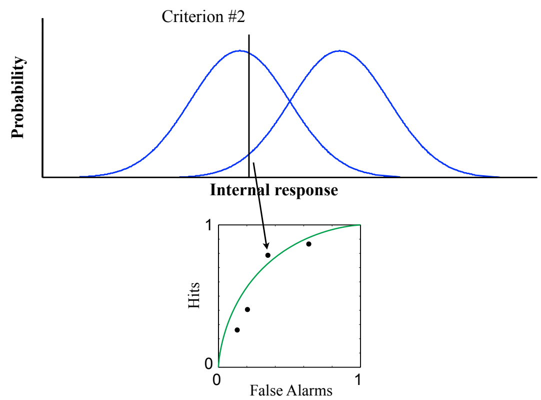
$$\beta = \frac{p(x=c\,|\,S+N)}{p(x=c\,|\,N)} = \frac{e^{-(c-d')^2/2}}{e^{-c^2/2}} \qquad \text{(Fix } \boldsymbol{\sigma} = 1\text{)}$$

## ROC (Receiver Operating Characteristic)

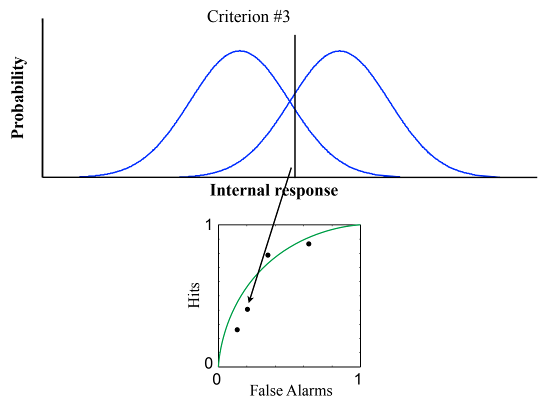



Plot anti-cumulatives:
*1-c(t|N)* vs. *1-c(t|S)*
as threshold t varies

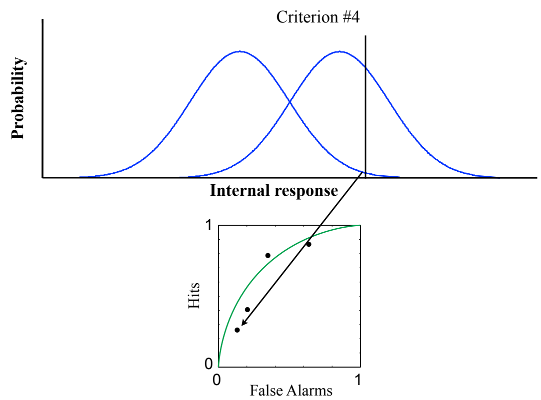
## ROC (Receiver Operating Characteristic)



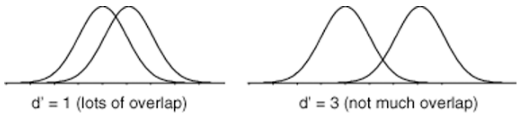

## ROC (Receiver Operating Characteristic)

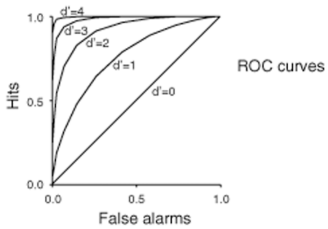



## ROC (Receiver Operating Characteristic)



## ROC (Receiver Operating Characteristic)



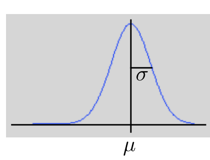

*[on board: Area under curve = %correct in a 2AFC task]*

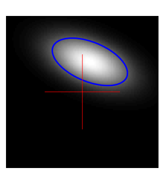
## Decision/classification in multiple dimensions

- Data-driven:
  - Fisher Linear Discriminant (FLD) - maximize d'
  - Support Vector Machine (SVM) - maximize margin
- Statistical:
  - ML/MAP/Bayes under a probabilistic model
  - e.g.: Gaussian, equal covariance (same as FLD)
  - e.g.: Gaussian, unequal covariance (QDA)
- Examples:
  - Visual gender identification
  - Neural population decoding

# Multi-D Gaussian densities

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



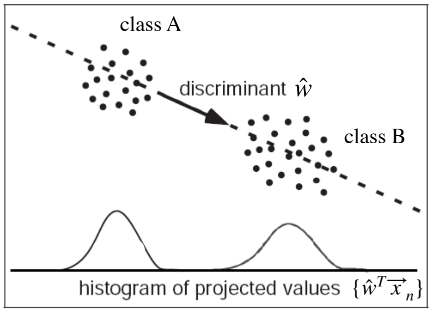



$$p(\vec{x}) = \frac{1}{\sqrt{(2\pi)^N |C|}}\, e^{-(\vec{x}-\vec{\mu})^T C^{-1} (\vec{x}-\vec{\mu})/2}$$

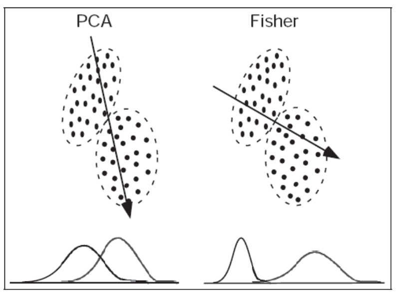
mean: [0.2, 0.8]
cov: [1.0 -0.3;
-0.3 0.4]

## Linear Classifier

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




## Fisher Linear Discriminant




$$\max_{\hat{w}} \frac{\left[\hat{w}^T(\mu_A - \mu_B)\right]^2}{\left[\hat{w}^T C_A \hat{w} + \hat{w}^T C_B \hat{w}\right]}$$

(note: this is "d-prime" !)

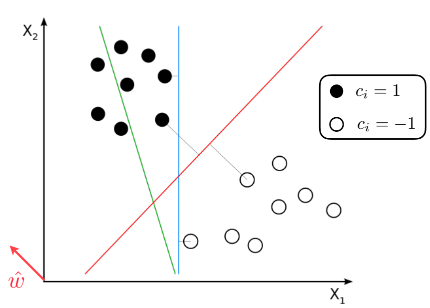
optimum: $\hat{w} = C^{-1}(\mu_A - \mu_B)$ , where $C = \frac{1}{2}(C_A + C_B)$

## Support Vector Machine (SVM)

(widely used in machine learning, has no closed form)

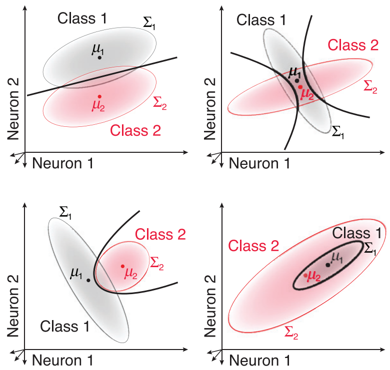
Maximize the "margin" (gap between data sets)

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



## ML (or MAP) classifier assuming Gaussians

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



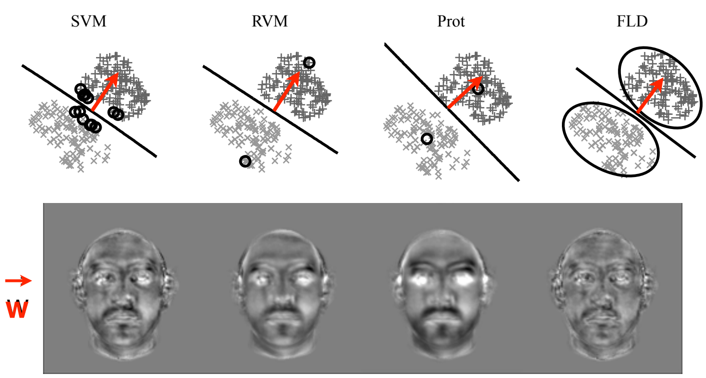

[figure: Pagan et al. 2016]

## Perceptual example: Gender identification

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

[Graf & Wichmann, NIPS*03]

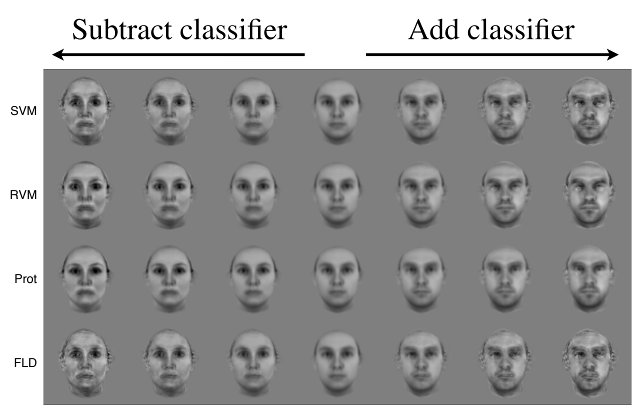
# Linear classifiers



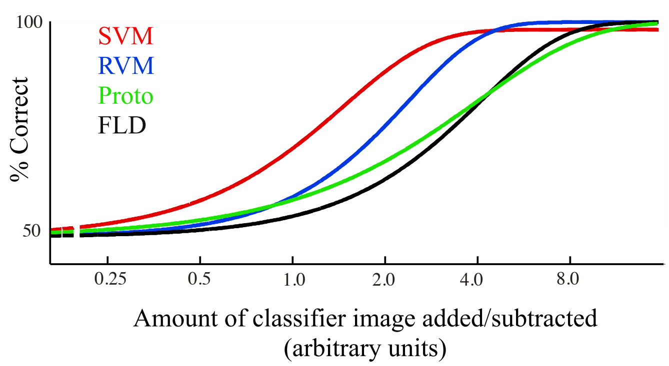

Four linear classifiers trained on subject data

# 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
- Curse of dimensionality strongly limits this result. A more direct test: Synthesize optimally discriminable faces...




[Wichmann, et. al; NIPS*04]



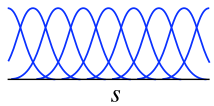

[Wichmann, et. al; NIPS*04]

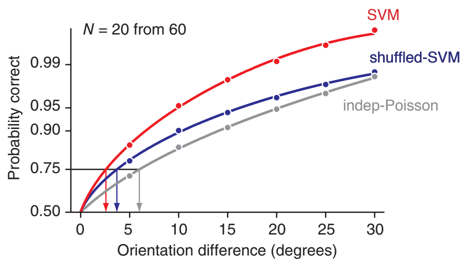
# Population decoding

Independent Poisson responses
[e.g., Seung & Sompolinsky, 1993]

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

Tuning curves, $h_n(s)$




[Graf, Kohn, Jazayeri, Movshon, 2011]