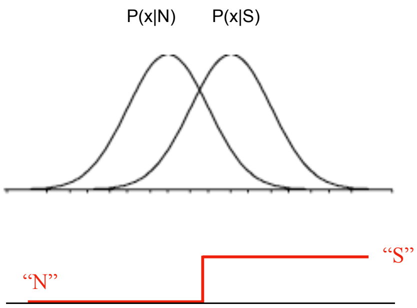
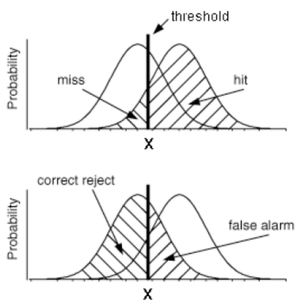
Signal Detection Theory



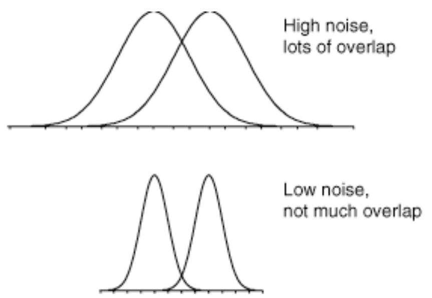
For equal, unimodal, symmetric distributions, ML decision rule is a *threshold.* Bayes decision rule is *shifted* threshold.

Signal Detection Theory: experimental outcomes

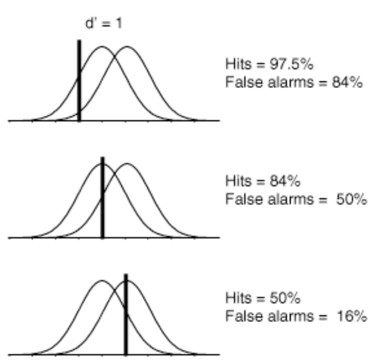


	"Yes"	"No"
Light Present	Hit	Miss
No Light Present	False Alarm	Correct Rejection

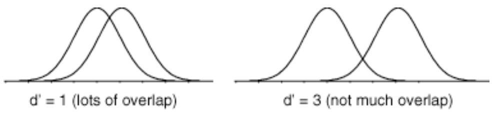
Signal Detection Theory: discriminability (d’)

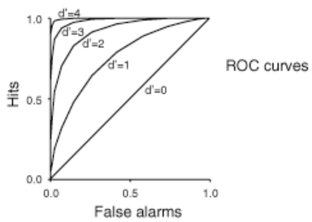


Signal Detection Theory: Criterion



Signal Detection Theory: “receiver operating characteristic” (ROC)





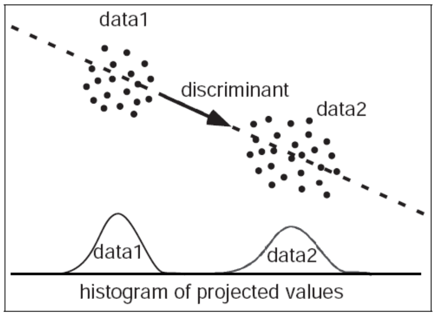
[on board: Area under curve = %correct in a 2AFC task]

Discriminants 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)
- Regularization
- Examples:
  - Visual gender identification
  - Neural population decoding

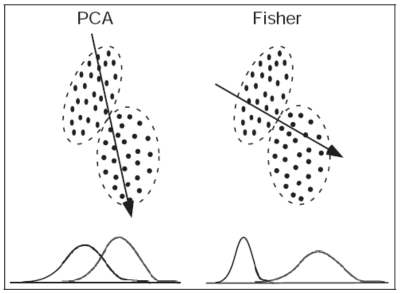
Linear Classifier

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



Simplest choice: difference of means

Fisher Linear Discriminant



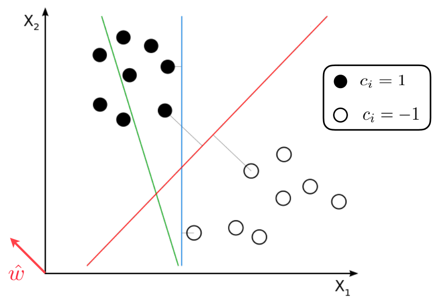
$$\max_{\hat{w}} \frac{\left[\hat{w}^T(\mu_A - \mu_B)\right]^2}{\left[\hat{w}^T C_A \hat{w} + \hat{w}^T C_B \hat{w}\right]}$$

$$\hat{w} = D^{-1} V^T (\mu_A - \mu_B), \quad \text{where} \quad V D^2 V^T = C_A + C_B$$

Support Vector Machine

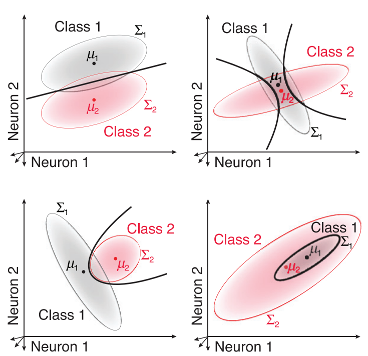
Maximize "margin" (gap between data sets)

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



Gaussian ML classifier

Linear (for equal covariances), or quadratic (three different geometries).



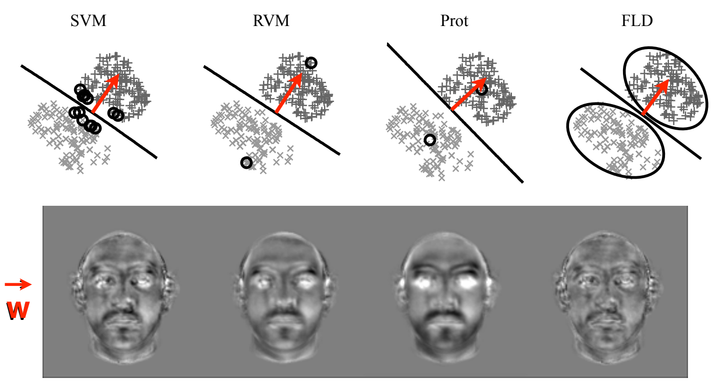
[figure: Pagan et al. 2016]

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



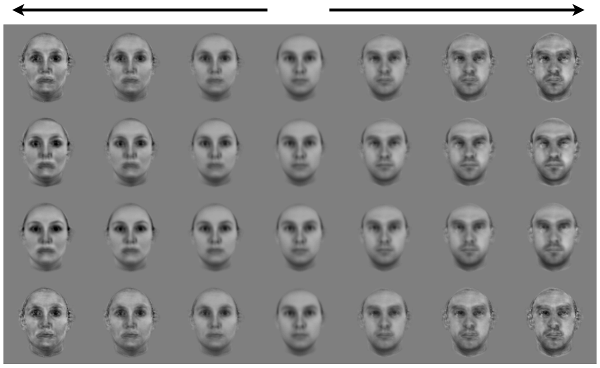
Four linear classifiers trained on subject data

Model validation/testing

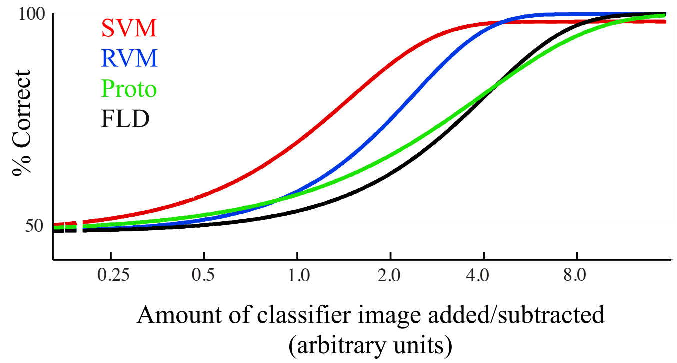
- Cross-validation: Subject responses [% correct, reaction time, confidence] are explained
  - very well by SVM
  - moderately well by RVM / FLD
  - not so well by Prot
- Curse of dimensionality strongly limits this result. A more direct test: Synthesize optimally discriminable faces...

Subtract classifier ← → Add classifier

SVM
RVM
Prot
FLD



[Wichmann, Graf, Simoncelli, Bülthoff, Schölkopf, NIPS*04]



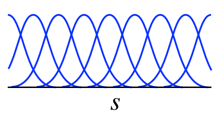
[Wichmann, Graf, Simoncelli, Bülthoff, Schölkopf, NIPS*04]

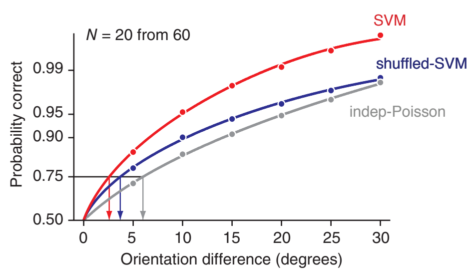
Population decoding

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, $h_n(s)$





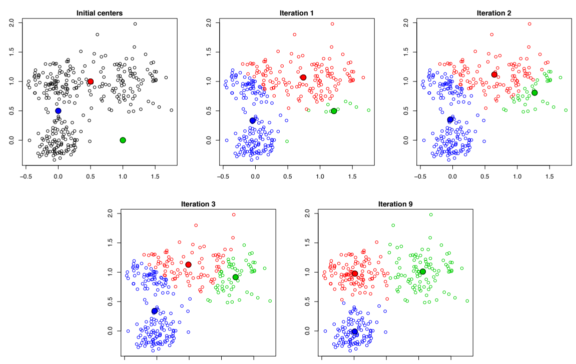
Graf, Kohn, Jazayeri, Movshon, 2011

Clustering

- K-Means
- ML, assuming Gaussians
  => Soft-assignment K-means
  (a form of Expectation-Maximization - EM)

K-means example

Here $X_i \in \mathbb{R}^2$, $n = 300$, and $K = 3$



[from R. Tibshirani, 2013]

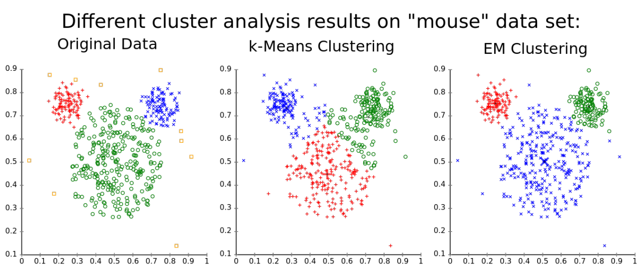
Mixture of Gaussians ML clustering

$$p(\vec{x}_n|a_{ni},\vec{\mu}_i,C_i) \propto \sum_i \frac{a_{ni}}{\sqrt{|C_i|}} e^{-(\vec{x}_n-\vec{\mu}_i)^T C_i^{-1}(\vec{x}_n-\vec{\mu}_i)/2}$$

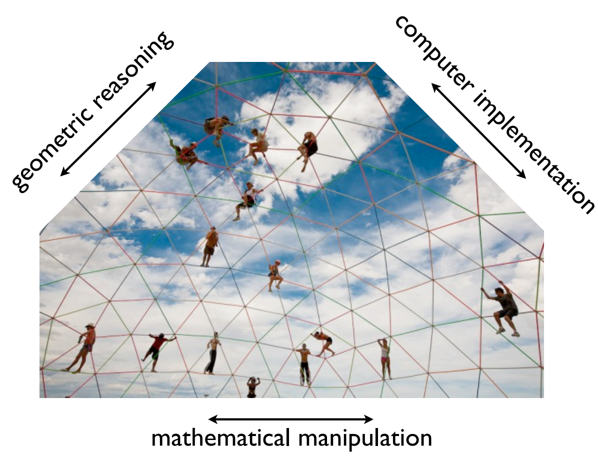
a_{ni} = assignment probability

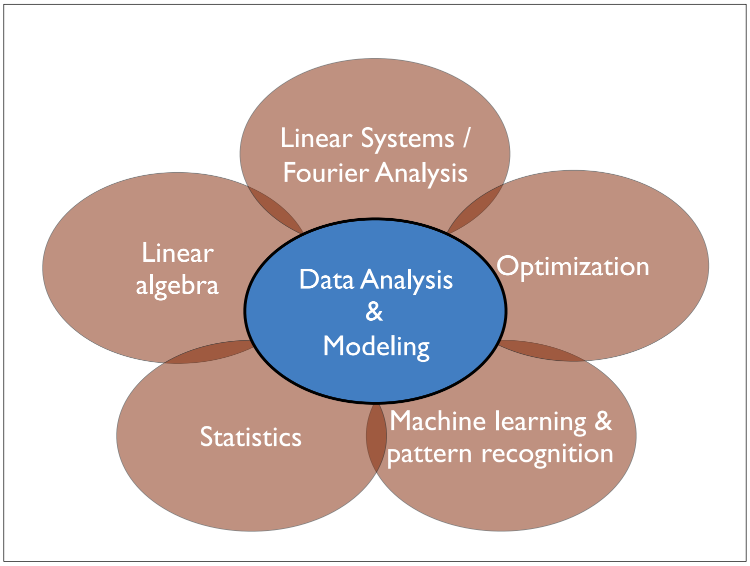
$\{\vec{\mu}_i, C_i\}$ = mean/covariance of class i

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



[wikipedia]


geometric reasoning
computer implementation
mathematical manipulation


Linear Systems /
Fourier Analysis
Linear
algebra
Optimization
Data Analysis
&
Modeling
Statistics
Machine learning &
pattern recognition