Pattern Classification and Recognition:

Bayes Classifiers

ECE 681

Spring 2016

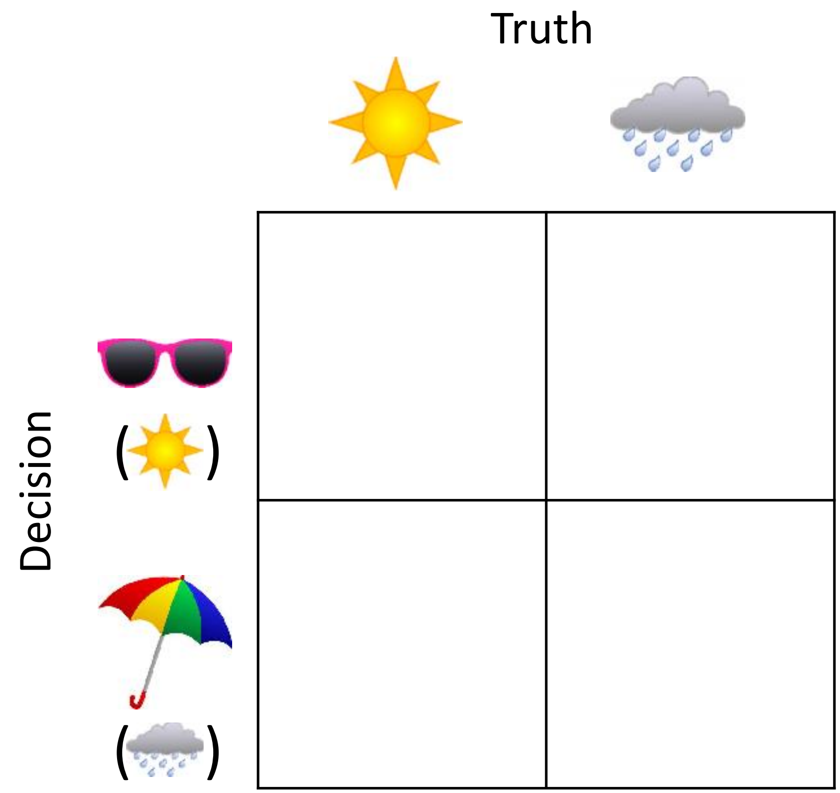
Stacy Tantum, Ph.D.

Bayesian Decision Theory

Quantifies trade-offs between decisions using

- Available data

- Probability
  - *prior*
  - *likelihood*
  - *evidence*
- Costs of decision outcomes



Bayes' Theorem

Rev. Thomas Bayes

Rev. Thomas Bayes presented a solution to the problem of *inverse probability*

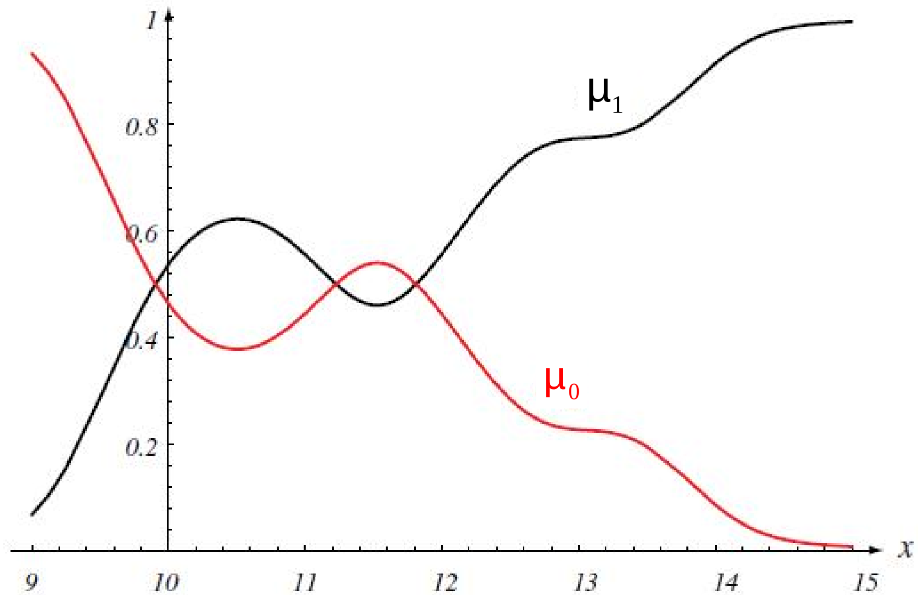
- Given data, what can we say about the process that generated it?
- What is the probability of the underlying state from which observed data was generated?

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

$$posterior = \frac{likelihood \times prior}{evidence}$$

Making Decisions

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



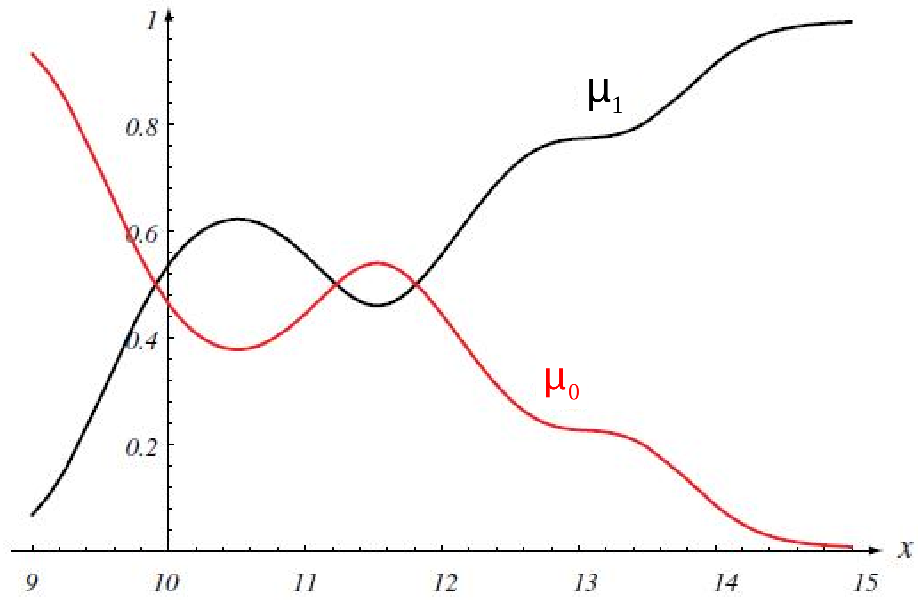
Making Decisions

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

Bayes Decision Rule (alternate form)

Decide μ_0 if $p(x \mid \mu_0)P(\mu_0) > p(x \mid \mu_1)P(\mu_1)$

Decide μ_1 otherwise



Making Cost-Aware Decisions

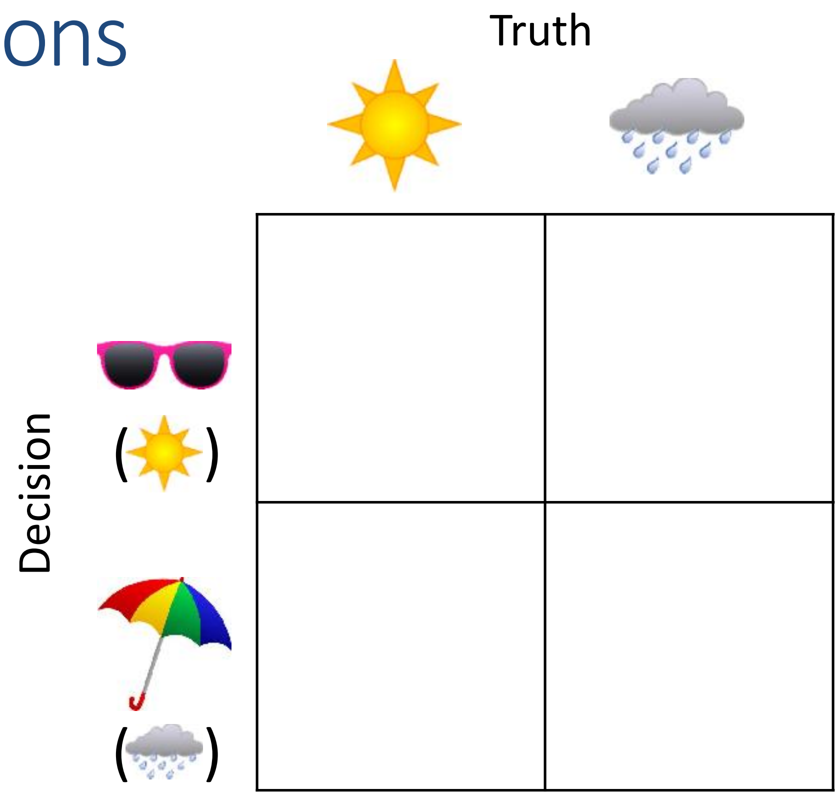
Every action (decision) has a cost, which depends on the "truth"

$$c_{ij} = \mathrm{cost}(\sim_i \mid \mu_j)$$

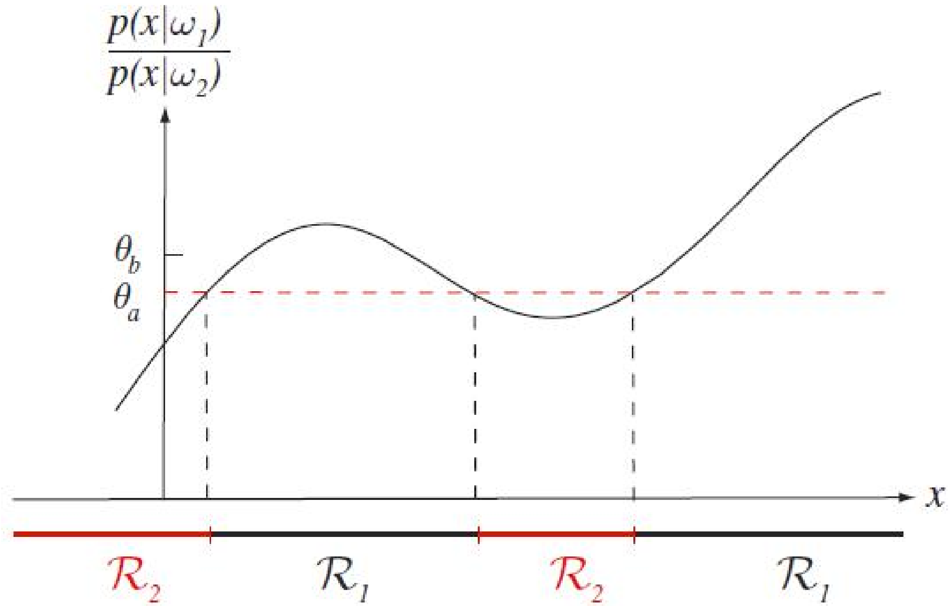
Risk of a decision = expected cost

$$R(\sim_0 \mid x) = \mathrm{c}_{00}\, \mathrm{P}(\mu_0 \mid \mathrm{x}) + \mathrm{c}_{01}\, P(\mu_1 \mid x)$$

$$R(\sim_1 \mid x) = \mathrm{c}_{10}\, \mathrm{P}(\mu_0 \mid \mathrm{x}) + \mathrm{c}_{11}\, P(\mu_1 \mid x)$$



Likelihood Ratio



Discriminant Functions

Generalized function of the data, $g_i(x)$, that supports *discriminating* class i from other candidate classes

$$\text{Decide } \mu_0 \text{ if } g_0(x) > g_1(x)$$
$$\text{Decide } \mu_1 \text{ if } g_1(x) > g_0(x)$$

For the general case with risks

$$g_i(x) = -R(\sim_i \mid x)$$

Two-Class Discriminant Functions

Combine $g_0(x)$ and $g_1(x)$ into a single function

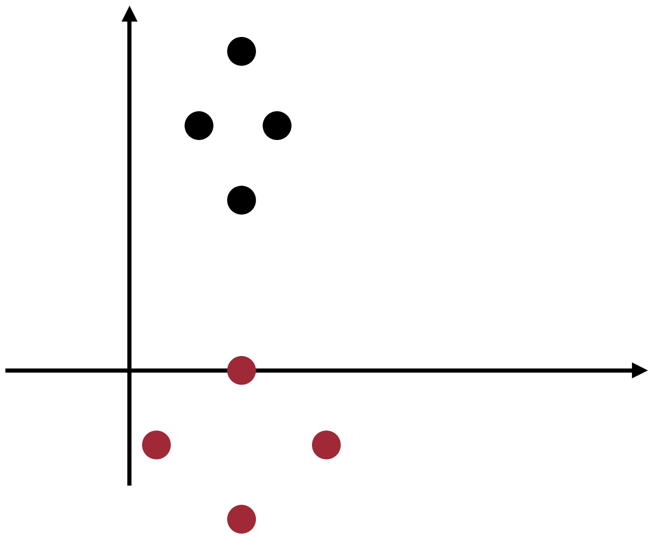
$$g(x) \equiv g_1(x) - g_0(x) \qquad\qquad \text{Decide } \mu_1 \text{ if } g(x) > 0$$

Discriminant Function for Normal Densities

$$p(\mathbf{x} \mid \mu_i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\circledR)^{d/2} \mid \boldsymbol{\Sigma}_i \mid^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right]$$

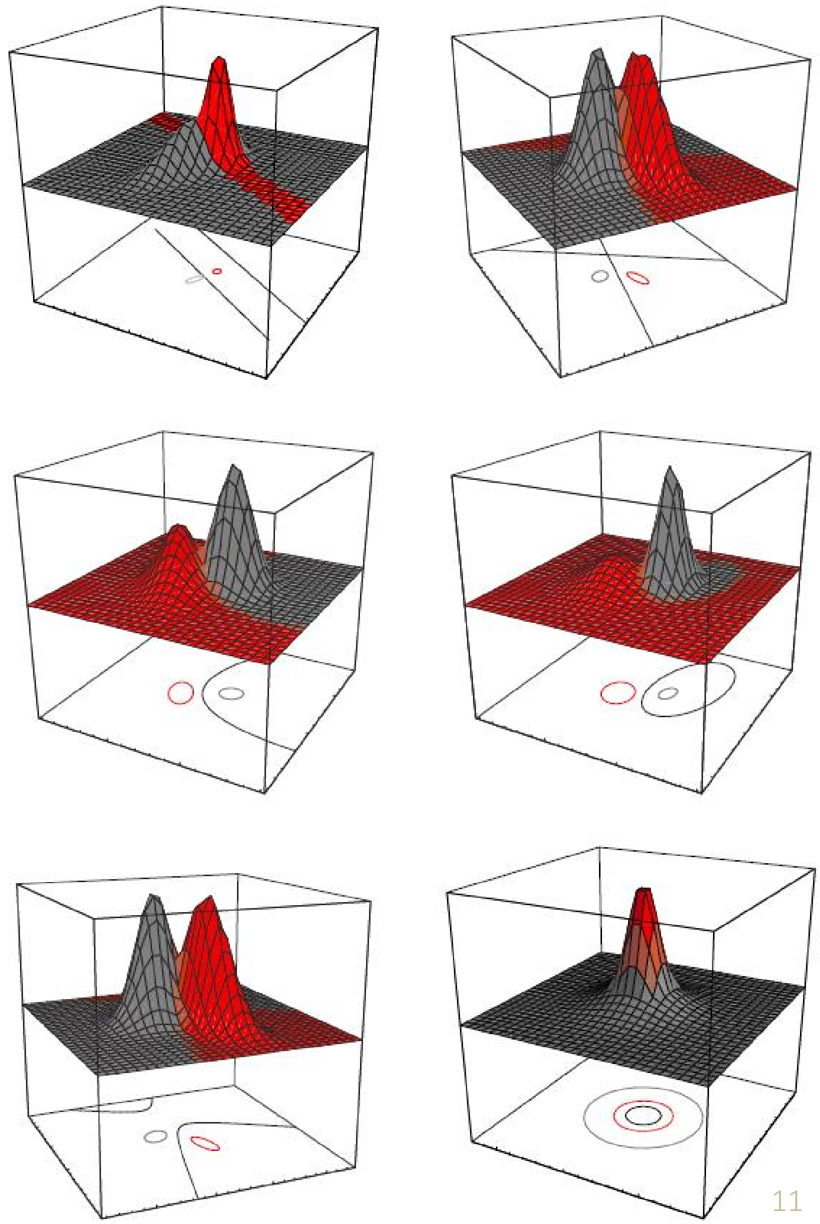
$$g_i(\mathbf{x}) = -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \frac{d}{2} \ln 2\circledR - \frac{1}{2} \ln \mid \boldsymbol{\Sigma}_i \mid + \ln P(\mu_i)$$

Decision Regions for Gaussian Data



$$\boldsymbol{\mu}_B = \begin{bmatrix} 3 \\ 6 \end{bmatrix} \qquad \boldsymbol{\Sigma}_B = \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\boldsymbol{\mu}_R = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \qquad \boldsymbol{\Sigma}_R = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$



Discrete Data

Data $\mathbf{x}$ takes on only 1 of m discrete values $\mathbf{v}_1$, $\mathbf{v}_2$, … $\mathbf{v}_m$

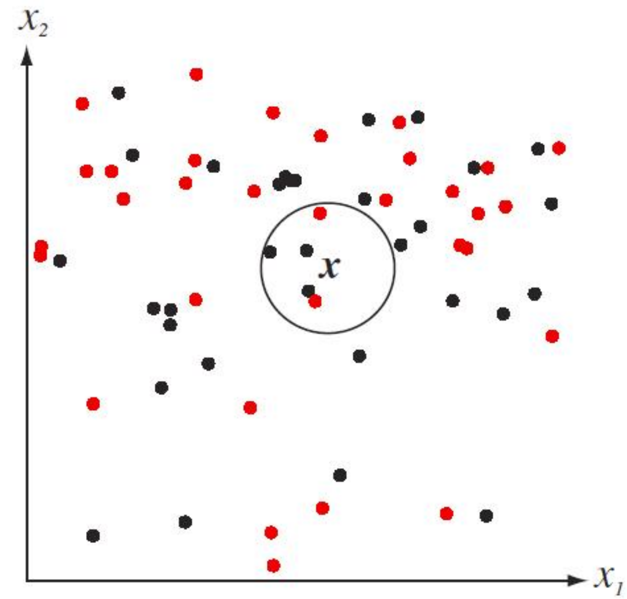
Binary Data

Data **x** takes on only 1 of *2* discrete values, i.e., 0 or 1

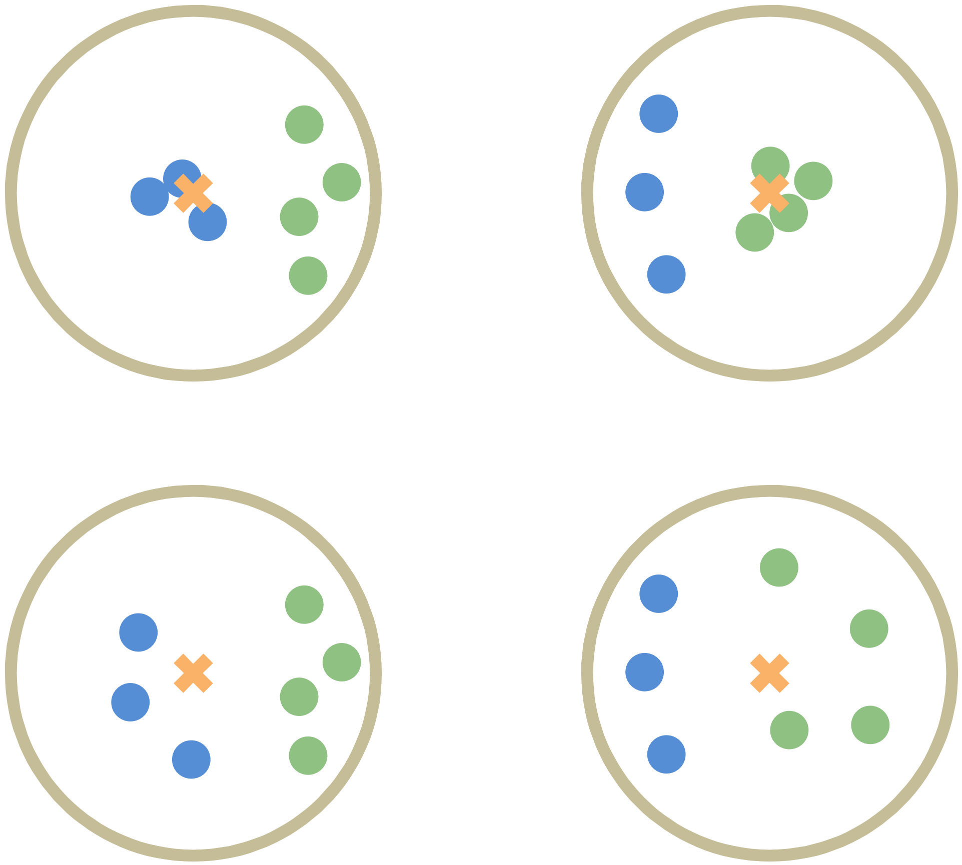
Relation to KNN

Numerical estimate of pdfs

- Total of n samples
- What proportion of the k samples in a volume V belong to class i?



KNN… Hmm…



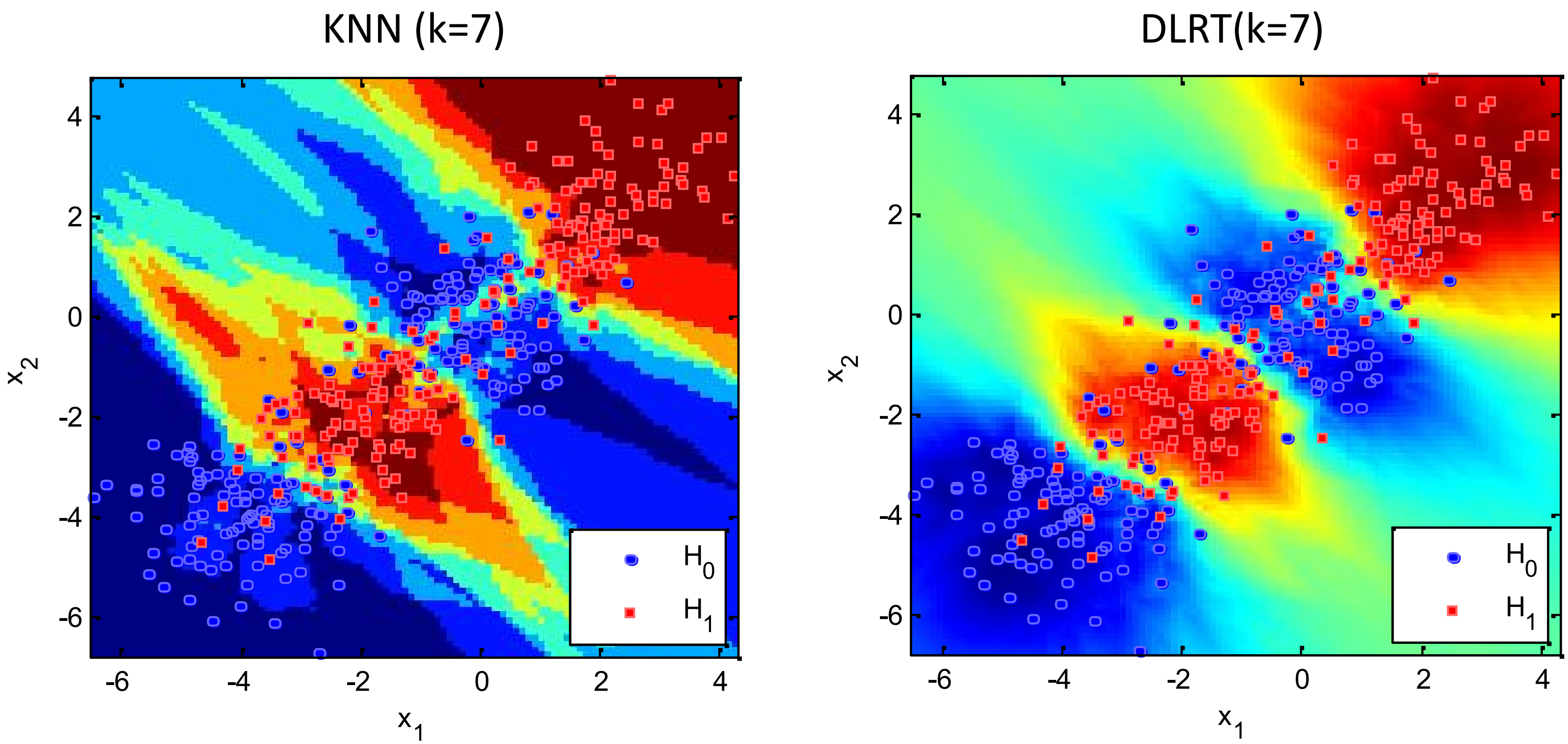
Distance Likelihood Ratio Test (DLRT)

Use KNN density estimation to *estimate* the likelihood ratio

- Considers distance to neighbors

J.J. Remus, K.D. Morton, P.A. Torrione, S.L. Tantum, and L.M. Collins,
``Comparison of a distance-based likelihood ratio test and k-nearest neighbor classification methods,''
in *Machine Learning for Signal Processing*, (Cancun, Mexico), pp. 362--367, 2008.

KNN and DLRT Comparison



Bayes (Gaussian pdfs)

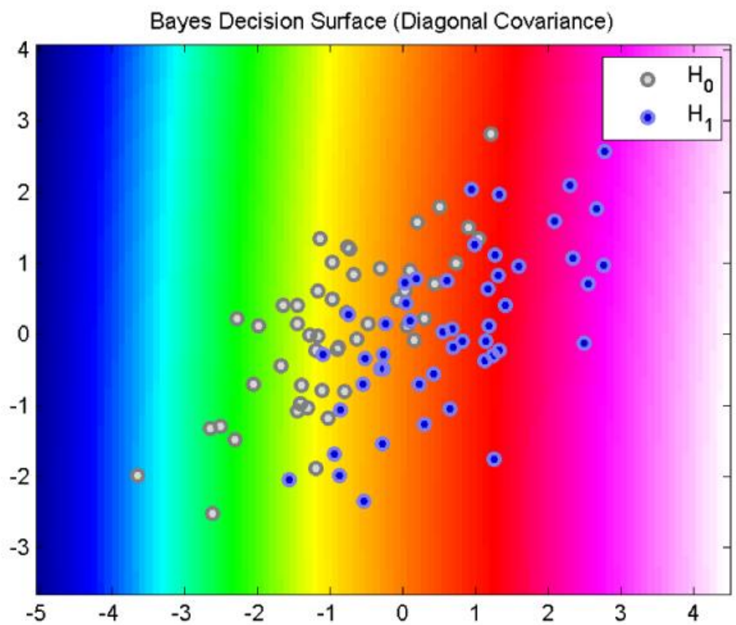
(Generalized) Likelihood Ratio Test

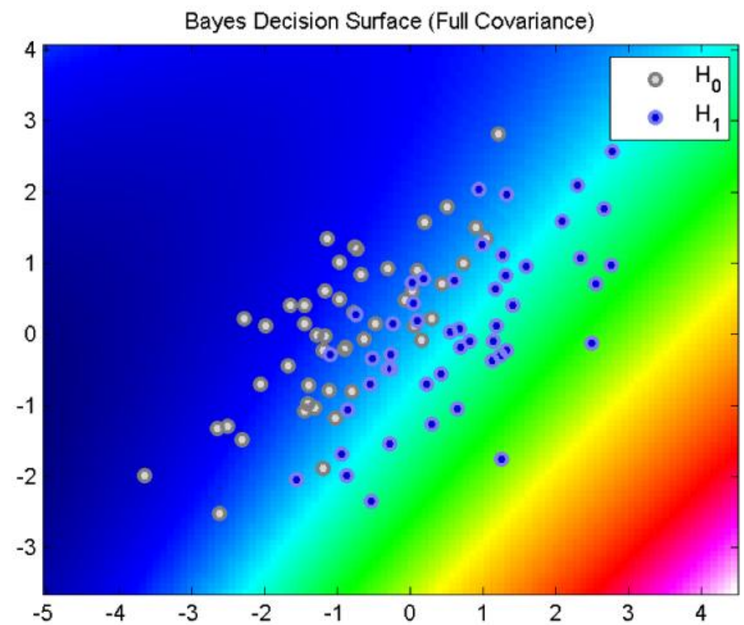
Decision Statistic: $g_1(x) - g_0(x)$

$$= \left[-\frac{1}{2}(x - \text{«}_1)^T \Sigma^{-1}(x - \text{«}_1) - \frac{1}{2}\ln|\Sigma_1| + \ln P(\mu_1) \right]$$

$$- \left[-\frac{1}{2}(x - \text{«}_0)^T \Sigma^{-1}(x - \text{«}_0) - \frac{1}{2}\ln|\Sigma_0| + \ln P(\mu_0) \right]$$

See Matlab functions `mean`, `cov`, and `diag`





Bayes (Gaussian pdfs)

Decision Statistic: $g_1(x) - g_0(x)$

$$= \left[-\frac{1}{2}(x - \text{«}_1)^T \Sigma^{-1}(x - \text{«}_1) - \frac{1}{2}\ln|\Sigma_1| + \ln P(\mu_1) \right]$$

$$- \left[-\frac{1}{2}(x - \text{«}_0)^T \Sigma^{-1}(x - \text{«}_0) - \frac{1}{2}\ln|\Sigma_0| + \ln P(\mu_0) \right]$$

Training a Bayes classifier:

- What do we need to run it?

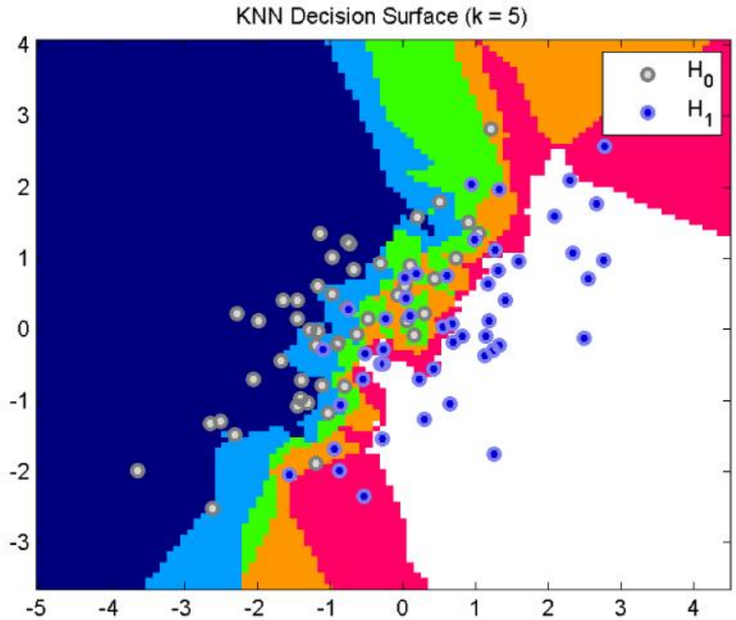
Running a Bayes classifier:

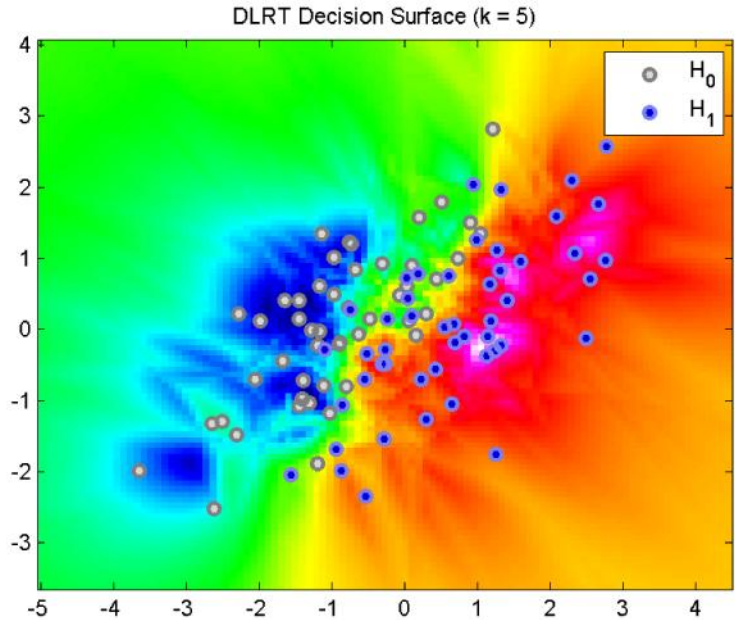
- How do we get a decision statistic?

DLRT

Decision Statistic: $\ln\left(\frac{n_0}{n_1}\right) + D\left[\ln \Delta_{k0} - \ln \Delta_{k1}\right]$

Like KNN, but takes into account the distance to the k^{th} neighbor





DLRT

Decision Statistic: $\ln\left(\frac{n_0}{n_1}\right) + D\left[\ln \Delta_{k0} - \ln \Delta_{k1}\right]$

See Matlab function `log` (for ln, `log10` is for $\log_{10}$)

Training a DLRT:

- What do we need to run it?

Running a DLRT:

- How do we get a decision statistic?