Recognition by finding patterns

- We have seen very simple template matching (under filters)
- Some objects behave like quite simple templates
  - Frontal faces

- Strategy:
  - Find image windows
  - Correct lighting
  - Pass them to a statistical test (a classifier) that accepts faces and rejects non-faces
- Important high level point:
  - Want to develop some understanding of the relationship of modelling statistics and deciding between options

Basic ideas in classifiers

- Loss
  - some errors may be more expensive than others
    - e.g. a fatal disease that is easily cured by a cheap medicine with no side-effects -> false positives in diagnosis are better than false negatives
  - We discuss two class classification: L(1->2) is the loss caused by calling 1 a 2
- Total risk of using classifier s

$$R(s) = Pr\left\{1 \to 2 | \text{using } s\right\} L(1 \to 2) + Pr\left\{2 \to 1 | \text{using } s\right\} L(2 \to 1)$$

Details of formula optional, but the idea is worth understanding

Basic ideas in classifiers

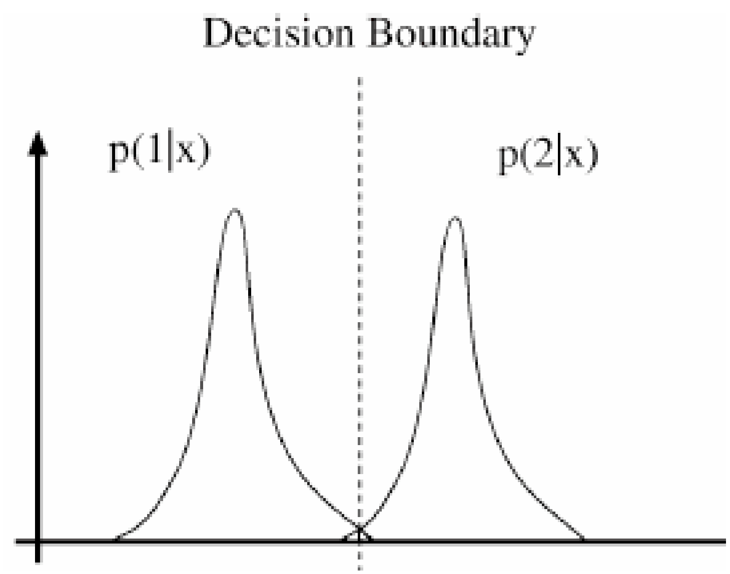
- Generally, we should classify as 1 if the expected loss of classifying as 1 is better than for 2
- gives

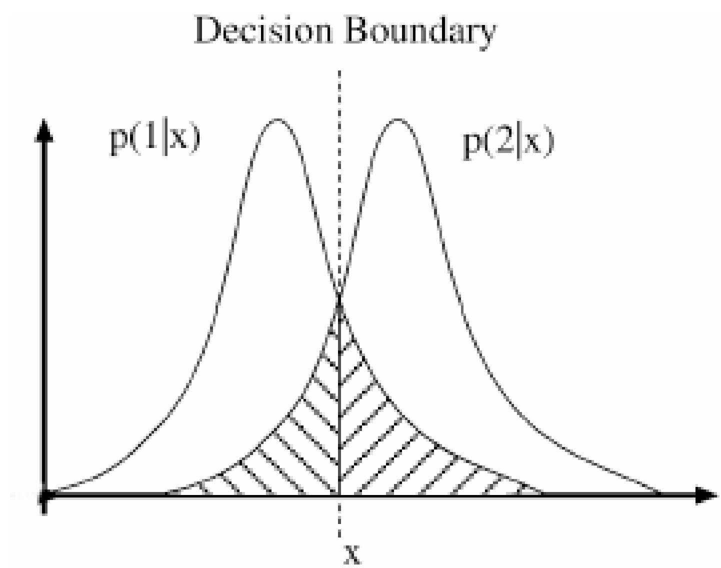
1 if $p(1|x)L(1 \to 2) > p(2|x)L(2 \to 1)$

2 if $p(1|x)L(1 \to 2) < p(2|x)L(2 \to 1)$

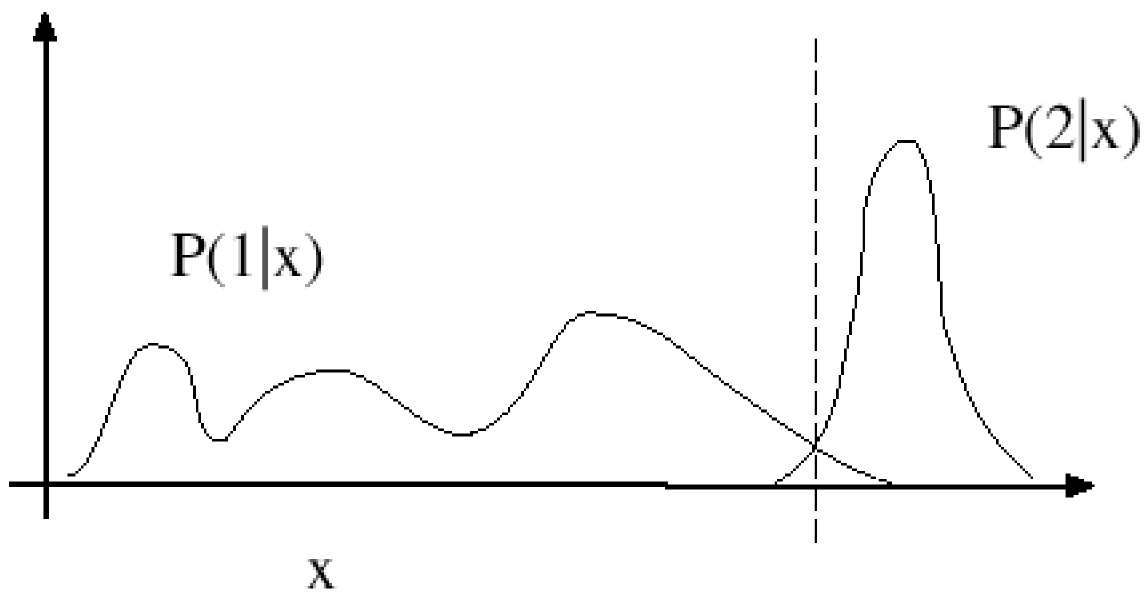
- Crucial notion: Decision boundary
  - points where the loss is the same for either case

Some loss may be inevitable: the minimum risk (function of the decision boundary) is the area of the shaded resion is called the Bayes risk

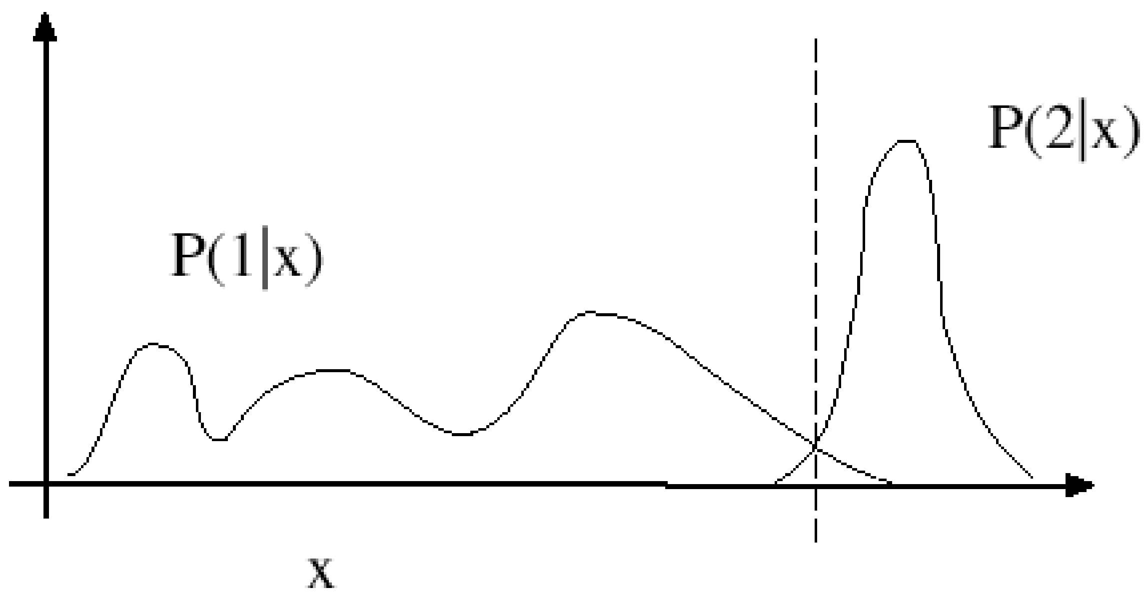




Finding a decision boundary is not the same as modelling a conditional density.



Finding a decision boundary is not the same as modelling a conditional density.



Important point: P(l|x) can be inaccurate, but the system can work well, as long as the boundary is correct)

Plug-in classifiers

- Assume that distributions have some parametric form - now estimate the parameters from the data.
- Typical example:
  - assume a normal distribution with shared covariance, different means; us usual estimates
  - ditto, but different covariances
- Issue: parameter estimates that are "good" may not give optimal classifiers.

Example: known distributions

$$p(\underline{x}|k) = \left(\frac{1}{2\pi}\right)^{-p/2} |\underline{\underline{\Sigma}}|^{-1/2} \exp\left[-\frac{1}{2}\left(\underline{x}-\underline{\mu}_k\right)^T \underline{\underline{\Sigma}}^{-1}\left(\underline{x}-\underline{\mu}_k\right)\right]$$

- Assume normal (Gaussian) class densities, multi-dimensional measurements with common (known) covariance and different (known) means
- Class priors are π_k
- Posteriors for class k given observation x is then:

$$p(k \mid x) \propto (\pi_k)\left(\frac{1}{2\pi}\right)^{-p/2} |\underline{\underline{\Sigma}}|^{-1/2} \exp\left[-\frac{1}{2}\left(\underline{x}-\underline{\mu}_k\right)^T \underline{\underline{\Sigma}}^{-1}\left(\underline{x}-\underline{\mu}_k\right)\right]$$

- Classifier boils down to:
  choose class that minimizes:

$$\Delta\left(\underline{x}, \underline{\mu}_k\right)^2 - 2\log \pi_k$$

where

Mahalanobis distance ——

$$\Delta\left(\underline{x}, \underline{\mu}_k\right) = \left[\left(\underline{x} - \underline{\mu}_k\right)^T \underline{\Sigma}^{-1}\left(\underline{x} - \underline{\mu}_k\right)\right]^{1/2}$$

because covariance is common, this simplifies to sign of a linear expression:

Histogram based classifiers

- Use a histogram to represent the class-conditional densities
  - (i.e. $p(x|1)$, $p(x|2)$, etc)
- Advantage: estimates become quite good with enough data!
- Disadvantage: Histogram becomes big with high dimension
  - One way to deal with this is to assume feature independence

Finding skin

- Skin has a very small range of (intensity independent) colours, and little texture
  - Compute an intensity-independent colour measure, check if colour is in this range, check if there is little texture (median filter)
  - See this as a classifier - we can set up the tests by hand, or learn them.
  - get class conditional densities (histograms), priors from data (counting)
- Classifier is

  - if $p(\text{skin}|\boldsymbol{x}) > \theta$, classify as skin
  - if $p(\text{skin}|\boldsymbol{x}) < \theta$, classify as not skin
  - if $p(\text{skin}|\boldsymbol{x}) = \theta$, choose classes uniformly and at random

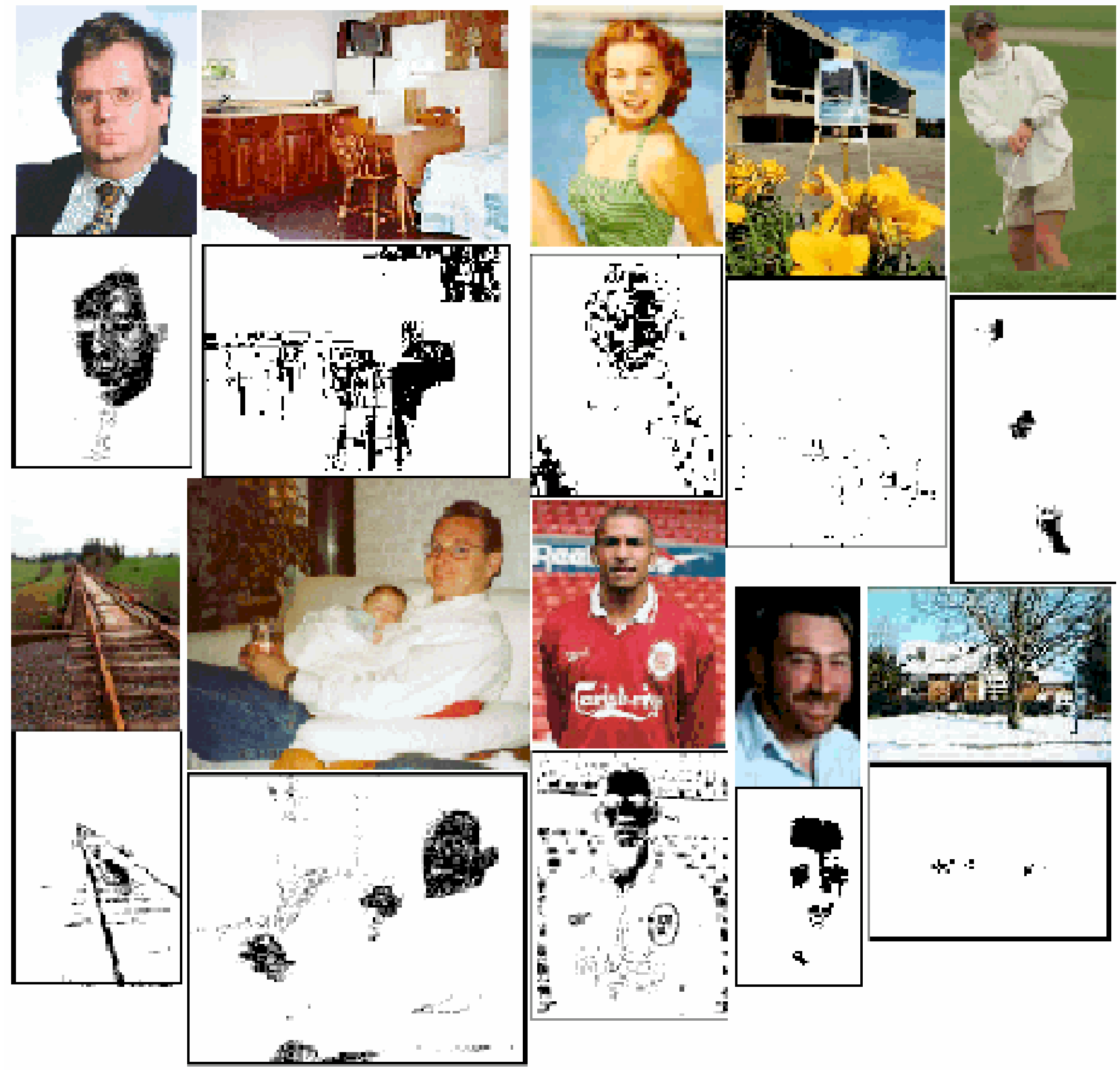


Figure from “Statistical color models with application to skin detection,” M.J. Jones and J. Rehg, Proc. Computer Vision and Pattern Recognition, 1999 copyright 1999, IEEE

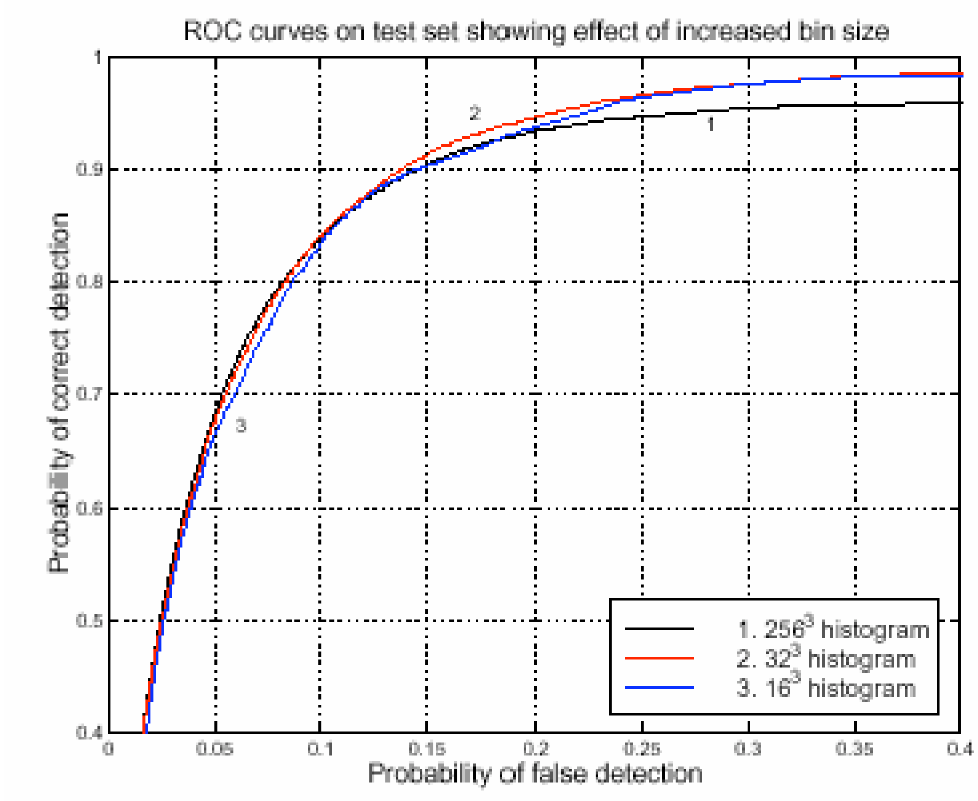


Figure from "Statistical color models with application to skin detection," M.J. Jones and J. Rehg, Proc. Computer Vision and Pattern Recognition, 1999 copyright 1999, IEEE