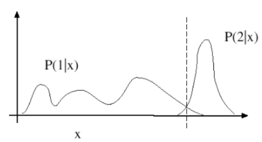
Building classifiers (continued)

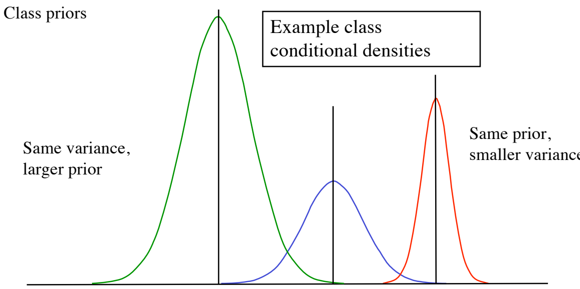
Now consider this one

- Standard scenario
  - Have training data
  - Want to classify new data
- One approach
  - Estimate the probability distributions (we have being thinking about them all along, e.g. P(1|x))
  - Issue: parameter estimates that are “good” may not give optimal classifiers
- Another approach
  - Directly go for the boundary



Example: known distributions (simple)

- Assume normal (Gaussian) class densities with different variances and means
- Two slightly non-obvious considerations for the classification boundary
  - Variances
  - Class priors



Example: known distributions (simple)

- Assume normal (Gaussian) class densities with different variances and means

$$p(x|k) = \left(\frac{1}{2\pi\sigma_k^2}\right)^{-1/2} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right)$$

- Class priors are $p(k) = \pi_k$
- Posteriors for class k given observation x is then (Bayes):

$$p(k|x) \propto p(k)p(x\mid k) = \pi_k \left(\frac{1}{2\pi\sigma_k^2}\right)^{-1/2} \exp\left[-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right]$$

Example: known distributions (full featured version)

Skipped in 2008

- Assume normal (Gaussian) class densities, multi-dimensional measurements with different covariances and means

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

- Class priors are π_k
- Posteriors for class k given observation x is then (Bayes):

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

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

Example --- 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

$$
\begin{aligned}
&\text{if } p(skin \mid \mathbf{x}) > \theta, \text{ then classify as skin} \\
&\text{if } p(skin \mid \mathbf{x}) < \theta, \text{ then classify as not skin} \\
&\text{if } p(skin \mid \mathbf{x}) = \theta, \text{ then choose at random}
\end{aligned}
$$

- Note that **x** is generally vector, unlike the simple sex from height example.

Classification by empirical distribution (summary)

- Bottom line---if we have enough training data (often do not!), then we estimate $P(c|x)$ by the empirical distribution (e.g., histogram of counts).
- May want to smooth the histogram (soft-binning)
  - Generally a very good idea---but can be expensive without good coding
- **Test by cross validation**

Optional in 2008

Naive Bayes

- The histogram classifier that estimates empirically $P(X_1, X_2, X_3, \ldots \mid c)$ (c == class)
- As the number of dimensions goes up, this gets out of hand
- Common but often effective approximation is to assume independence
  - $P(X_1, X_2, X_3, \ldots \mid c) = P(X_1 \mid c)P(X_2 \mid c)P(X_3 \mid c) \ldots$
  - (Naive Bayes)

Matching by relations

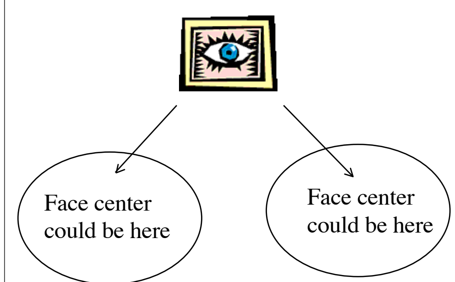
- Idea:
  - find bits, then say object is present if relation between bits is OK
- Advantage:
  - objects with complex configuration spaces don't make good templates
    - internal degrees of freedom (people)
    - increasing complexity means template is more restrictive
    - occlusion
    - aspect changes
    - (possibly) shading
    - variations in color / texture

Simplest

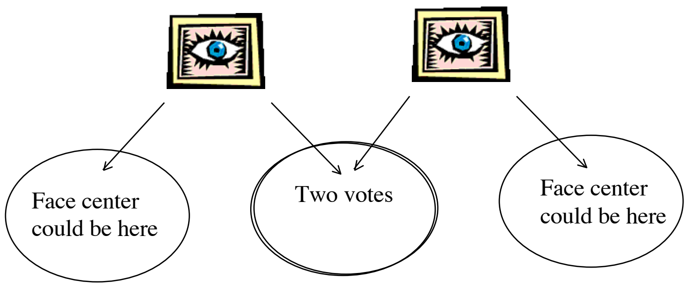
- Define a set of local feature templates
  - could find these with filters, etc.
  - corner detector+filters
- Think of objects as patterns
- Each template votes for all patterns that contain it
- Pattern with the most votes wins

Example: Faces

- Suppose that we specify the location of a face by the location of the nose.
- Further assume that left and right eyes are indistinguishable
- Further assume that we are looking for upright faces of specific size



(Assuming upright, and rough size)



(Assuming upright, and rough size)

Example: Faces

- Notice that if a part is missing (occluded, or simply missed by the detector), this can still work!
- What about the assumptions of size and orientation?
  - Brute force (try a bunch, perhaps taking cues from parts)
  - Some kind of strategy based on the prior
    - look for upright first, which is most likely to succeed if that does not work, try less common
  - The parts can vote for configuration and size
    - Voting array gets more dimensions

Example: Animals (e.g. humans)

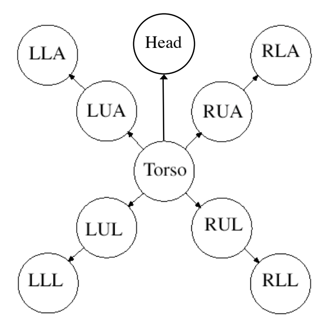
- Frontal faces are easy.
  - Configuration is relatively fixed
  - Number of key parts is not that big
- Articulated objects mean that the relation among parts can change (with priors)

People are trees

Technically, this means that the child parts only depend on parent parts

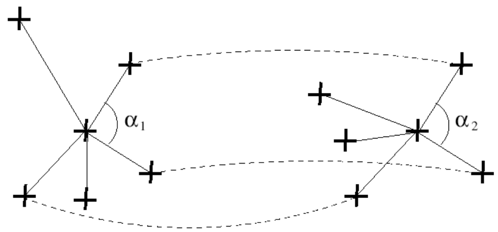
For example, the position (and orientation) of your hand depends on your lower arm. Knowing where the foot is does not give more information.

Clearly not true in general (examples?)



LLA=lower left arm
RUL=right upper leg

Employing spatial relations



a database entry and
its p closest features

a match

Figure from "Local gray value invariants for image retrieval," by C. Schmid and R. Mohr, IEEE Trans. Pattern Analysis and Machine Intelligence, 1997 copyright 1997, IEEE

Spatial Context Models

- Recognition often needs context!
- Objects posterior probability need to be a function of the object features, what is around, and what is known about the image and world in general
  - Context can be interpreted as a prior
- However, you also need to be able to recognize the tiger in the living room