Important

Cross-validation

- Split data set into two pieces, fit to one, and compute negative log-likelihood on the **other**
- One set is "training data", the other is "testing data" or "held out data"
- Average over different splits
- This estimates the quality of your model
  - Often (rightfully so) used to compare algorithms
- If you are doing model selection, then you choose the model with the smallest value of this average
  - This works because adding parameters causes over fitting of the training data which gives worse performance on test data

Skipped in 2008

Model averaging

- Often smarter to use multiple models for prediction than just one
- Consider that we have various models that we believe to various degrees, denoted by $P(M_i)$
- Suppose we want to estimate X from data, D, via the group of models, M_i
- A Bayesian would compute

$$P(X \mid D) = \sum_i P(X \mid M_i, D) P(M_i \mid D)$$

Recognition by finding patterns

- Template matching with correlation (linear filters) is a simple example of recognition by pattern matching
- Some objects behave like quite simple templates
  - Frontal faces

Recognition by finding patterns

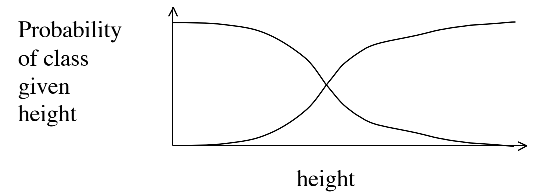
- Example strategy:
  - Find image windows
  - Correct for lighting
  - Pass them to a statistical test (a classifier) that accepts faces and rejects non-faces
- Important high level point:
  - Need to understand relationship between **modeling statistics** and deciding between options (classification AND risk analysis).

Basic ideas in classification

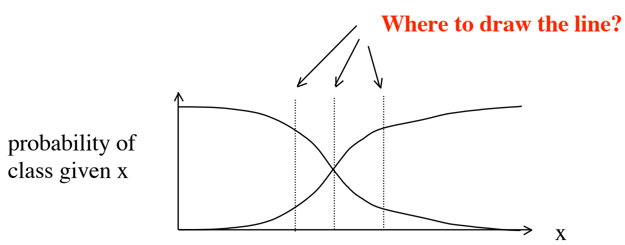
- Concrete example
  - "guess" male / female from height

Basic ideas in classification

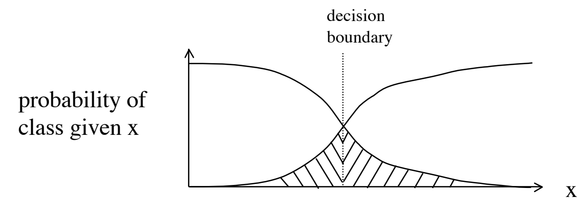
- Concrete example
  - "guess" male / female from height
- Probabalistic approach
  - Consider P(female|height)



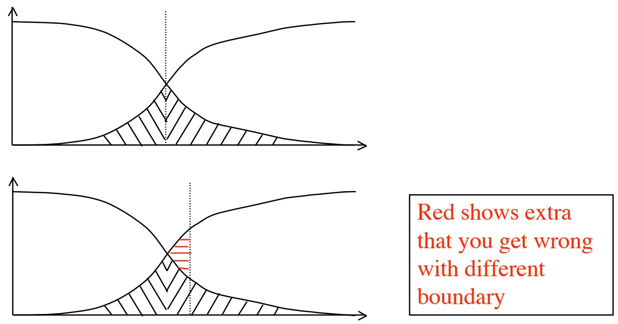
Basic ideas in classification



Basic ideas in classification



Area of intersection under curves gives expected value of making a mistake



Basic ideas in classification

- Concrete example
  - "guess" male / female from height
- Probabalistic approach
  - Consider P(female|height)
- Now consider "risk"
  - Suppose you want to give vaccine based on height for a disease that only males get.
  - There is great benefit to males who may be exposed
  - Vaccines have risk as well as benefit
  - Thus there is also some risk to giving females a vaccine they do not need
- How does this change the boundary?

Loss / Risk

- Some errors may be more expensive than others
  - e.g. a fatal disease that is easily cured by a cheap medicine with minimal side-effects --> false positives in diagnosis are better than false negatives
- We want to set the classification point
- Consider two class classification
  - Let L(1->2) be the loss caused by calling a 1 a 2
  - Want to anaylize the **expected value** of the loss (risk)

Basic ideas in classification

- Expected loss (risk) of using classifier *s*

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

Details of formula optional, but the idea is worth understanding

Basic ideas in classification

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

$$1 \text{ if } \quad \Pr(1|x)L(1 \to 2) > \Pr(2|x)L(2 \to 1)$$

$$2 \text{ if } \quad \Pr(2|x)L(2 \to 1) > \Pr(1|x)L(1 \to 2)$$

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

Details of formula optional, but the idea is worth understanding

Basic ideas in classification

- Expected loss (risk) of using classifier s

$$R(s) = L(1 \to 2) \bullet P(1 \to 2)(s) + L(2 \to 1) \bullet P(2 \to 1)(s)$$

Skipped in 2008

Basic ideas in classification

- Expected loss (risk) of using classifier s

$$R(s) = L(1 \to 2) \bullet P(1 \to 2)(s) + L(2 \to 1) \bullet P(2 \to 1)(s)$$

- So, given the class conditional densities, how do we set the boundary?

Skipped in 2008

Where is the boundary?

- Expected loss (risk) of using classifier s

$$R(s) = L(1 \to 2) \bullet P(1 \to 2)(s) + L(2 \to 1) \bullet P(2 \to 1)(s)$$

- Suppose that s is now our decision boundary, so x<s ==> 1; x>s ==> 2
- So, what is P(1->2)(s) ?

Skipped in 2008

Where is the boundary?

- Expected loss (risk) of using classifier *s*

$$R(s) = L(1 \to 2) \bullet P(1 \to 2)(s) \ + \ L(2 \to 1) \bullet P(2 \to 1)(s)$$

- Suppose that *s* is now our decision boundary, so x<s ==> 1; x>s ==> 2
- So, what is P(1->2)(s) ?

$$P(1->2)(s) = \int_s^{\infty} p(1,x)dx$$

Probability that we are a 1 in the region that we declare 2

Small p() reminds us that this is a probability density function, not a true probability. We use either P() or Pr() to empasize that we have a true probability. Further confusion arises because probability and probability density are very close if the domain is discrete (e.g. two classes).

Skipped in 2008

Where is the boundary?

- Expected loss (risk) of using classifier *s*

$$R(s) = L(1 \to 2) \bullet P(1 \to 2)(s) \ + \ L(2 \to 1) \bullet P(2 \to 1)(s)$$

- Suppose that *s* is now our decision boundary, so x<s ==> 1; x>s ==> 2
- So, what is P(1->2)(s) ?

$$P(1->2)(s) = \int_s^{\infty} p(1,x)dx$$

Probability that we are a 1 in the region that we declare 2

- Similarly,

Skipped in 2008

Where is the boundary?

- Expected loss (risk) of using classifier *s*

$$R(s) = L(1 \to 2) \bullet P(1 \to 2)(s) \ + \ L(2 \to 1) \bullet P(2 \to 1)(s)$$

- Suppose that *s* is now our decision boundary, so x<s ==> 1; x>s ==> 2
- So, what is P(1->2)(s) ?

$$P(1->2)(s) = \int_s^{\infty} p(1,x)dx$$

Probability that we are a 1 in the region that we declare 2

- Similarly,

$$P(2->1)(s) = \int_{-\infty}^{s} p(2,x)dx$$

Probability that we are a 2 in the region that we declare 1

Skipped in 2008

Where is the boundary?

- Expected loss (risk) is then

$$R(s) \ = \ \int_s^{\infty} L(1->2)\,p(1,x)dx \ + \int_0^{s} L(2->1)\,p(2,x)dx$$

- We want to minimize this

Skipped in 2008

Where is the boundary?

- Expected loss (risk) is then

$$R(s) = \int_s^{\infty} L(1->2)p(1,x)dx + \int_0^{s} L(2->1)p(2,x)dx$$

- We want to minimize this. So differentiate and set to 0.

$$\frac{d}{ds}\int_s^{\infty} L(1->2)p(1,x)dx = -L(1->2)p(1,s)$$

$$\frac{d}{ds}\int_{-\infty}^{s} L(2->1)p(2,x)dx = L(2->1)p(2,s)$$

(follows from the definition of integration and differentiation)

Skipped in 2008

Where is the boundary?

- Expected loss (risk) is then

$$R(s) = \int_s^{\infty} L(1->2)p(1,x)dx + \int_0^{s} L(2->1)p(2,x)dx$$

- We want to minimize this. Using previous pieces:

$$\frac{d}{ds}R(s) = L(2->1)p(2,s) - L(1->2)p(1,s)$$

- Setting to zero reveals the boundary for minimal risk:

$$L(2->1)p(2,s) = L(1->2)p(1,s)$$

Skipped in 2008

Basic ideas in classification

- Put differently

$$\begin{array}{ll} 1 \text{ if} & P(1|\mathrm{x})L(1\rightarrow 2) > \mathrm{P}(2|\mathrm{x})L(2\rightarrow 1) \\ 2 \text{ if} & P(2|\mathrm{x})L(2\rightarrow 1) > \mathrm{P}(1|\mathrm{x})L(1\rightarrow 2) \end{array}$$

- (Switching to conditional probability is OK here)
- Crucial intuitive notion: Decision boundary is at the points where the loss is the same for either case