CS4801 : Bayes classifier

Sahely Bhadra

21/8/2017

2. Bayes decision rule
   1. Classification error
   2. Minimum error rate classification
      1. Two category
      2. Multi category

Recap on classifier

A. No information

Random classifier

For k class classification problem assign class label k to a text point with probability 1/k

B. Prior or class probability is known

$$P(\text{error}) = \begin{cases} P(C_1) \text{ if decide } C_2 \\ P(C_2) \text{ if decide } C_1 \end{cases}$$

$$P(\text{error}) = \min \{ P(C_1), P(C_2) \}$$

Recap on classifier

A. No information

Random classifier

For k class classification problem assign class label k to a text point with probability 1/k

$$\mathbf{P(error)} = \begin{cases} P(C_1) \text{ if decide } C_2 \\ P(C_2) \text{ if decide } C_1 \end{cases}$$

B. Prior or class probability is known

$$\mathbf{P(error)} = \min \{ P(C_1), P(C_2) \}$$

C. Posterior is known

Bayes classifier

The Bayes rule is optimum, that is, it minimises the average probability error!

$$P(error/x) = \begin{cases} P(C_1/x_t) \text{ if decide } C_2 \\ P(C_2/x_t) \text{ if decide } C_1 \end{cases}$$

$$P(error/x) = \min \{ P(C_1/x_t), P(C_2/x_t) \}$$

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error/x)p(x)dx$$

Expected Loss

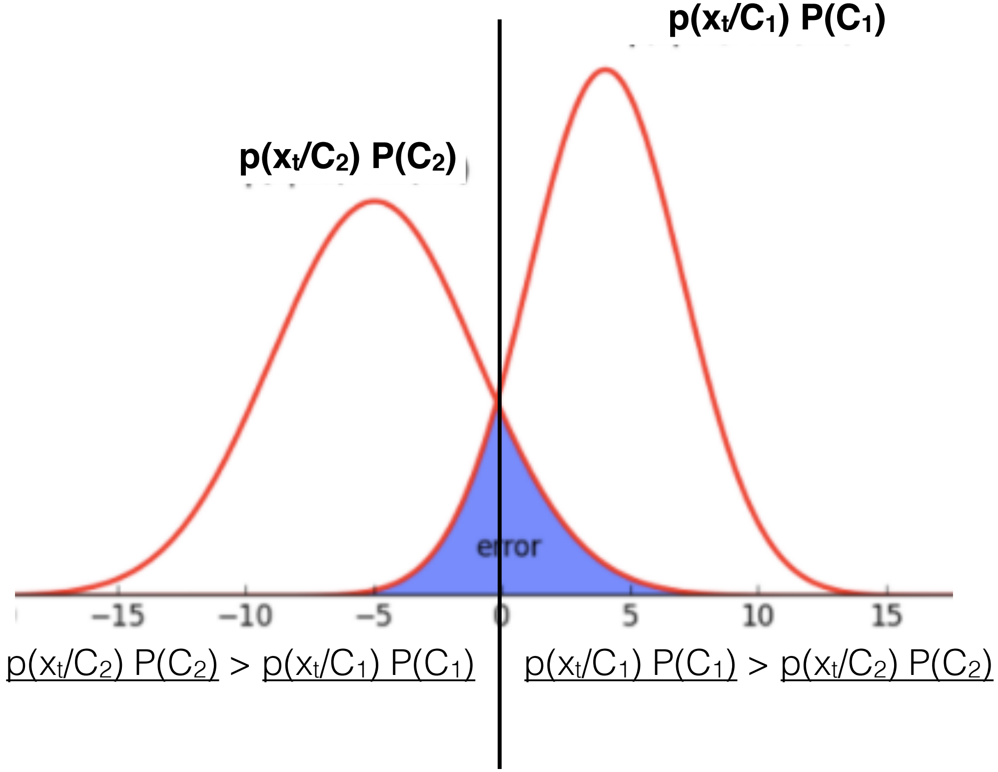
- Classification = f(x) : x -> $\{C_1, C_2\}$
- cost(f(x) , y)= **L(f(x) , y)** =cost of assigning class label 'f(x)' when the correct label is 'y'.

Expected loss

$$Risk(f) = E[R(x)] = \int_x R(x)\, p(x)\, dx$$

$$= \int_{f(x)=C_1} L(f(x),C_2) p(x/C_2)\, P(C_2)\, p(x)\, dx$$

$$+ \int_{f(x)=C_2} L(f(x),C_1) p(x/C_1)\, P(C_1)\, p(x)\, dx$$



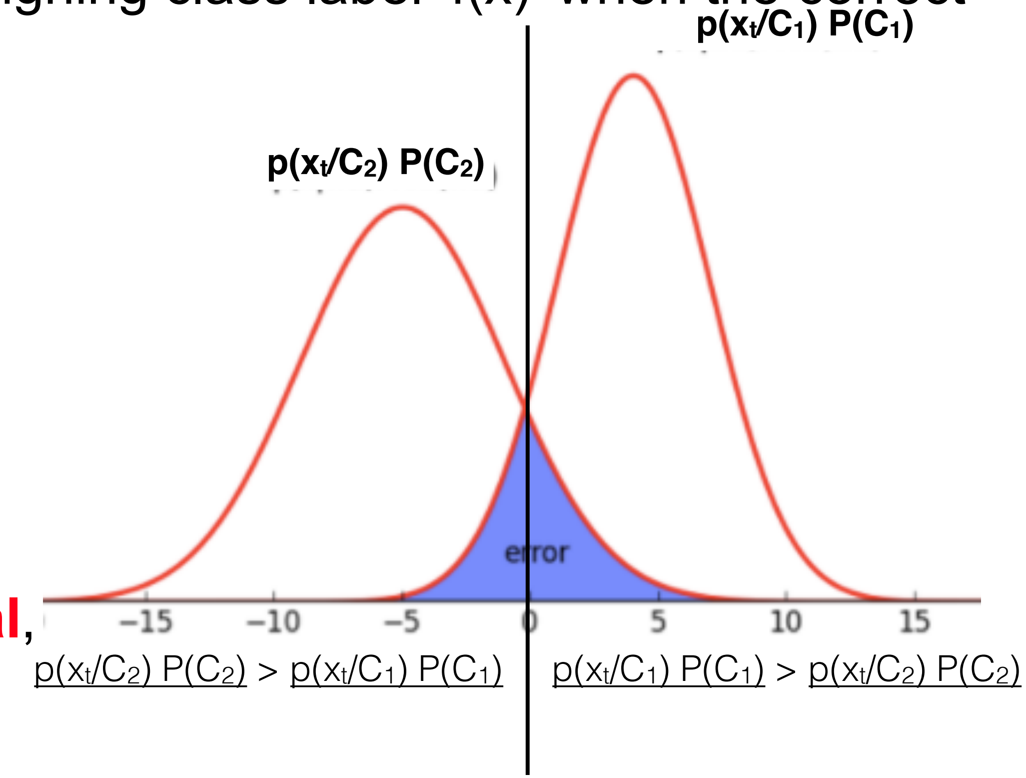
$$\boxed{Risk(f) = E[L(f(x),y)] = \int_{xy} L(f(x),y)\, p(x,y)\, d(x,y)}$$

Bayes Error

- Classification = $f(x) : x \rightarrow \{C_1, C_2\}$
- cost(f(x) , y)= **L(f(x) , y)** =cost of assigning class label 'f(x)' when the correct label is 'y'.

Expected loss or risk of classifier 'f'

$$\text{Risk}(f) = E[\,L(f(x),y)\,]$$
$$= \int_{xy} L(f(x),y)\,p(x,y)\,d(x,y)$$



A classifier f^* is called **Bayes optimal**, or **Bayes classifier**, if it minimises Risk(f).

Decide The minimum expected loss Risk(f^*) is called the **Bayes error**.

Bayes classifier

Classification = f(x) : x -> $\{C_1, C_2\}$
cost(f(x) , y)= **L(f(x) , y)** =cost of assigning class label 'f(x)' when the correct label is 'y'.

Risk(f) = E[L(f(x),y)]

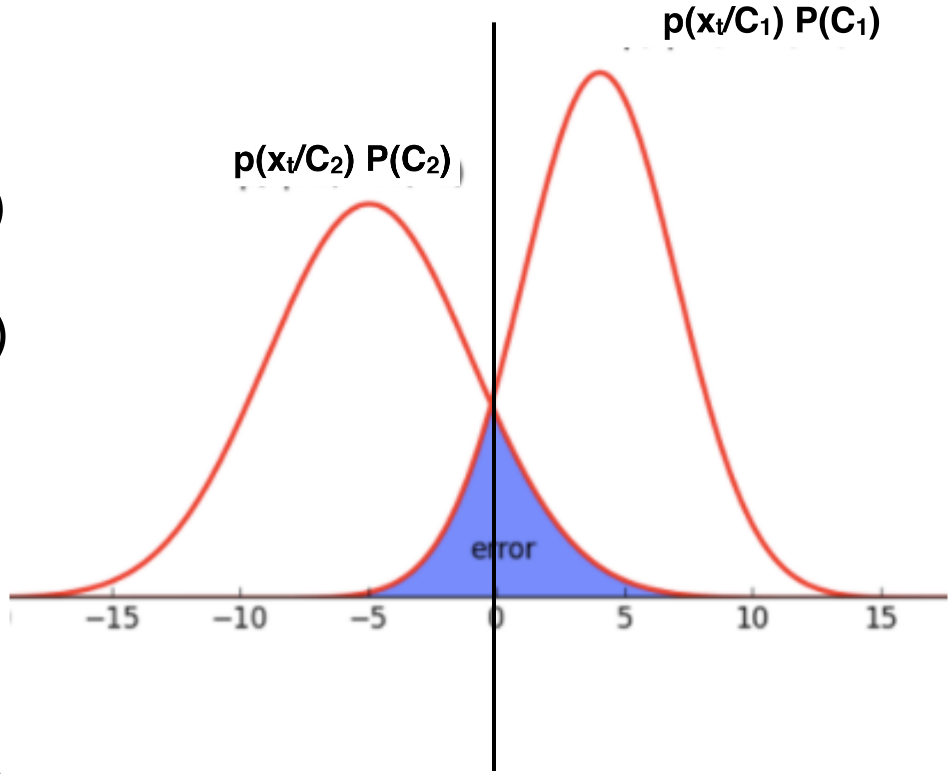
$Risk(f=C_1/x) = L_{11} P(C_1/x_t) + L_{12} P(C_2/x_t)$

$Risk(f=C_2/x) = L_{21} P(C_1/x_t) + L_{22} P(C_2/x_t)$

Bayes classifier : minimises Risk(f)

Decide $f=C_1$
if

$$Risk(f=C_1/x) < Risk(f=C_2/x)$$
$$(L_{11} - L_{21}) P(C_1/x_t) < (L_{22} - L_{12}) P(C_2/x_t)$$



Decide The minimum expected loss Risk(f*) is called the **Bayes error**.

Bayes classifier : Zero One Loss

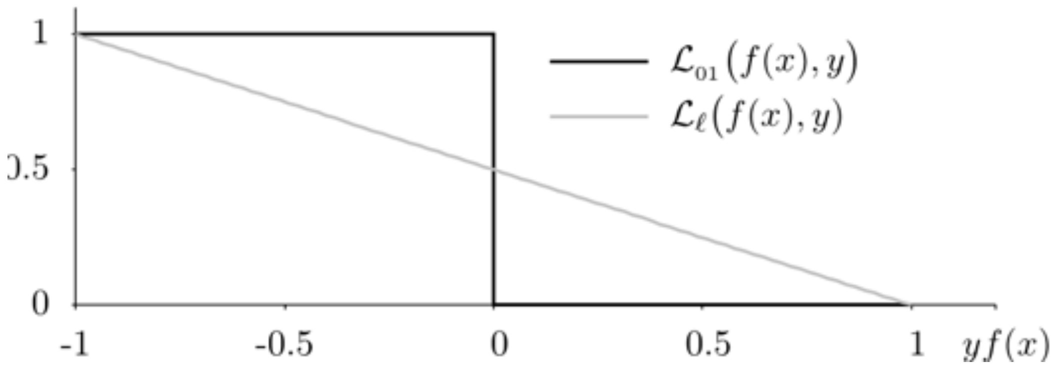
Bayes classifier : minimises Risk(f)

Decide $f=C_1$ if $(L_{11} - L_{21})\, P(C_1/x_t) < (L_{22} - L_{12})\, P(C_2/x_t)$ otherwise $f=C_2$

For Zero One Loss

$L_{11} = L_{22} = 0$

$L_{12} = L_{21} = 1$



Decide $f=C_1$

if $(0 - 1)\, P(C_1/x_t) < (0 - 1)\, P(C_2/x_t)$ or $P(C_2/x_t) < P(C_1/x_t)$

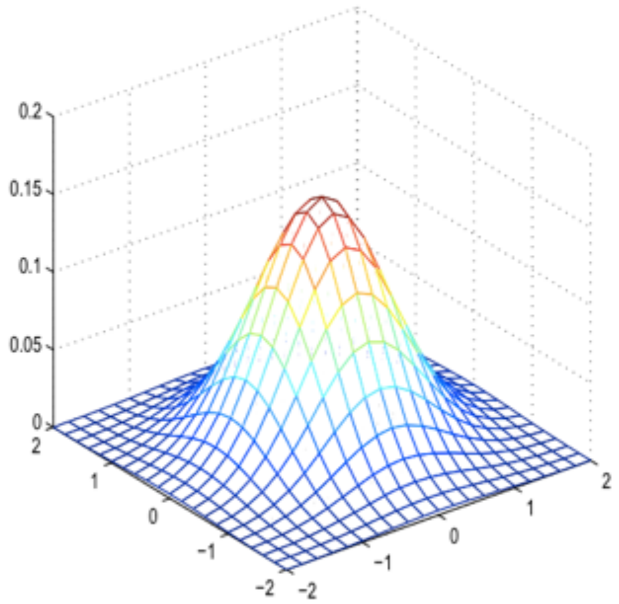
otherwise $f=C_2$

Naive Implementation

Naive Bayes

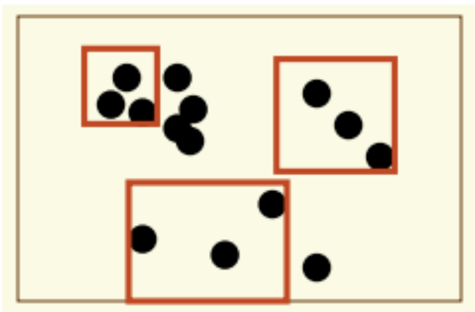
- Features are independent

Non parametric density estimation



$$P = \int_{\mathcal{R}} p(\mathbf{x}) d\mathbf{x} \approx p(\mathbf{x}) \int_{\mathcal{R}} d\mathbf{x} = p(\mathbf{x}) V$$

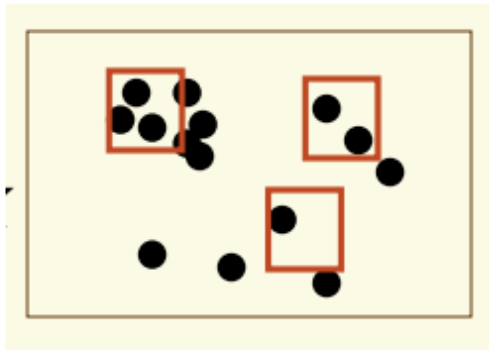
Nearest Neighbour (fixed k)



$$P = k/n$$

$$p(\mathbf{x}) = \frac{k/n}{V}$$

Parzen window (Fixed V)



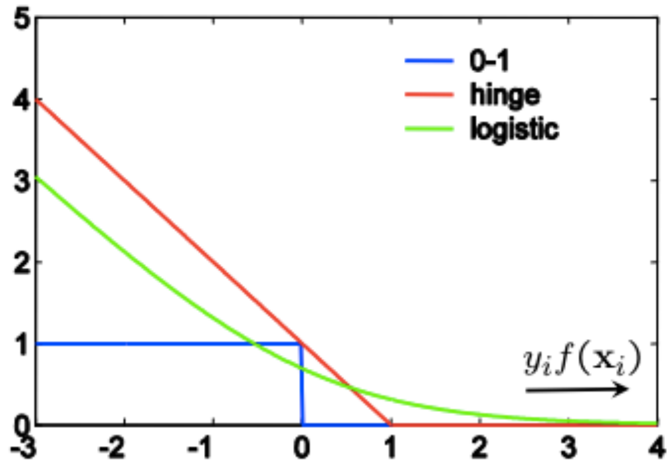
$$P = k/n$$

$$p(\mathbf{x}) = \frac{k/n}{V}$$

Classification Loss

Zero one loss
Logistic loss
Hinged loss

Square loss ?



Next Class

- 23/8
    - Logistic Regression
    - Perceptron classifier