IRFAN SHEIKH
16
MCSC 201

1) (a) Two classes, C_1 and C_2

Three actions, α_1 : choose C_1
α_2 : Choose C_2
α_R : reject

Let λ_{ik} = loss incurred for taking action α_i when the input actually belongs to C_k.

Loss matrix

action(Predicted) \ Actual	C_1	C_2
C1	0	0.8
C2	1	0
reject	0.5	0.5

The expected risk for taking action α_i given instance x,

$$R(\alpha_i|x) = \sum_{k=1}^{2} \lambda_{ik} P(C_k|x)$$

So,

$$R(\alpha_1|x) = \sum_{k=1}^{2} \lambda_{1k} P(C_k|x)$$

$$= \lambda_{11} P(C_1|x) + \lambda_{12} P(C_2|x)$$

$$= 0 + 0.8 P(C_2|x)$$

$$= 0.8 P(C_2|x)$$

$$R(\alpha_2|x) = \sum_{k=1}^{2} \lambda_{2k} P(C_k|x)$$

$$= \lambda_{21} P(C_1|x) + \lambda_{22} P(C_2|x)$$

$$= P(C_1|x) + 0 = P(C_1|x)$$

$$R(\alpha_R|x) = 0.5 P(C_1|x) + 0.5 P(C_2|x)$$

$$= 0.5 (P(C_1|x) + P(C_2|x))$$

$$= 0.5$$

Optimal Decision Rules

- choose action α_i if $R(\alpha_i|x) < R(\alpha_k|x)$ for all $k \neq i$
  and $R(\alpha_i|x) < R(\alpha_R|x)$
- reject if $R(\alpha_R|x) < R(\alpha_i|x)$, $i = 1, 2, \ldots$

Now,
$$R(\alpha_1|x) = 0.8\, P(C_2|x)$$
$$R(\alpha_2|x) = P(C_1|x)$$
$$R(\alpha_R|x) = 0.5$$

action α_1 will be chosen if
$$0.8\, P(C_2|x) < P(C_1|x)$$
and $0.8\, P(C_2|x) < 0.5$

$\Rightarrow$ $0.8\, P(C_2|x) < 1 - P(C_2|x)$; $P(C_2|x) < \frac{0.5}{0.8}$

$\Rightarrow$ $1.8\, P(C_2|x) < 1$

$\Rightarrow$ $P(C_2|x) < \frac{1}{1.8}$

$\Rightarrow$ $P(C_2|x) < \frac{10}{18}$ and $P(C_2|x) < \frac{5}{8}$

$\Rightarrow$ $P(C_2|x) < \frac{10}{18}$

and $\cancel{P(C_1|x)}$

$1 - P(C_1|x) < \frac{10}{18}$

$\Rightarrow$ $P(C_1|x) > \frac{8}{10}$ $\quad$ ∴ option A

$= 0.8$

Choose action α_2 if

$$P(C_1|x) > 0.8\, P(C_2|x) \quad \text{and } P(C_1|x) > 0.5$$

$$= 0.8(1 - P(C_1|x))$$

$$= 0.8 - 0.8P(C_1|x)$$

$$\Rightarrow 1.8\, P(C_1|x) > 0.8$$

$$\Rightarrow P(C_1|x) > \frac{8}{18} \quad \text{and} \quad P(C_1|x) > 0.5$$

$\Rightarrow$ choose α_2 when $P(C_1|x) > 0.5$

Choose action α_R if

$$0.5 < P(C_1|x) \text{ and } 0.5 < P(C_2|x)$$

2.) parameters (α, β^2)

data sample X, with N instances.

log likelihood, $\mathcal{L}(\alpha, \beta|X) = -N\log\beta - \frac{\sum_t (x^t - \alpha)^2}{2\beta^2}$

<u>MLE for α</u>

Taking partial derivative wrt α and setting it to 0, we get.

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \frac{-1}{2\beta^2}\, 2\sum_t (x^t - \alpha)\cdot(-1) = 0$$

$$\Rightarrow \frac{1}{\beta^2}\sum_t (x^t - \alpha) = 0$$

$$\Rightarrow \sum_t (x^t - \alpha) = 0$$

$$\Rightarrow \sum_{t=1}^{N} x^t = \sum_{t=1}^{N} \alpha$$

$$\Rightarrow \frac{\sum_{t=1}^{N} x^t}{N} = \alpha$$

Thus, the MLE estimate for α is $\frac{\sum_{t=1}^{N} x^t}{N}$.

MLE for β

Taking partial derivative wrt β and setting it to 0, we get:

$$\frac{\partial L}{\partial \beta} = -\frac{N}{\beta} - \frac{\sum_t (x^t-\alpha)^2}{2} - (2\sigma^{-3}) = 0$$

$$\Rightarrow \frac{N}{\beta} = \frac{\sum_{t=1}^{N} (x^t-\alpha)^2}{\beta^3}$$

$$\Rightarrow \frac{N\beta^3}{\beta} = \sum_{t=1}^{N} (x^t-\alpha)^2$$

$$\Rightarrow \beta^2 = \frac{\sum_{t=1}^{N} (x^t-\alpha)^2}{N}$$

where we use the MLE estimate for α; $\frac{\sum_t x^t}{N}$ above.

3.) Sample $X = \{x^t, r^t\}$
model $g(\cdot)$ trained over it.

Noise : is the part of error that can never be removed. It explains the variance of response r^t given an instance x^t, i.e. on the same x^t, how much r^t varies.

Bias : explains how much the expected response $E[g(x^t)]$ varies from the actual response r^t, disregarding the effect of varying samples.

Variance : explains that when we go from one sample to another, on average, how much $g(x^t)$ varies around the expected value, $E[g(x^t)]$.