

G V V Sharma*

CONTENTS

1	Mean Square Error	1
2	Bayes Classifier	1
3	Exercises	2

Abstract—This manual provides an introduction to statistical decision theory.

1 MEAN SQUARE ERROR

1.1 Let (\mathbf{X}, Y) be an input/output dataset, whose relation is unknown. If

$$\hat{Y} = f(\mathbf{X}) \quad (1.1)$$

be the estimate of Y , the *loss* in the estimate is defined as

$$L(Y, \hat{Y}) = L(f) \quad (1.2)$$

1.2 Let

$$L(f) = (Y - \hat{Y})^2 \quad (1.3)$$

Then

$$E[L(f)] \quad (1.4)$$

is defined as the *mean square error* (MSE). Show that

$$E[L(f)] = E_{\mathbf{X}} \{E_Y [Y - f(\mathbf{x})]^2 | \mathbf{X} = \mathbf{x}\} \quad (1.5)$$

1.3 Let

$$c = f(\mathbf{x}) \quad (1.6)$$

Using (1.5)

$$\begin{aligned} \min E[L(f)] &= \min L(f) | \mathbf{X} \\ &= \min_c E_Y \{ [Y - c]^2 | \mathbf{X} = \mathbf{x} \} \quad (1.7) \end{aligned}$$

Show that

$$\begin{aligned} L(f) | \mathbf{X} &= E_Y \{ [Y - c]^2 | \mathbf{X} = \mathbf{x} \} \\ &= -2c E_Y \{ Y | \mathbf{X} = \mathbf{x} \} + E_Y \{ Y^2 | \mathbf{X} = \mathbf{x} \} + c^2 \quad (1.8) \end{aligned}$$

1.4 $L(f)$ is minimum when

$$\frac{d}{dc} L(f) | \mathbf{X} = 0. \quad (1.9)$$

Show that this results in

$$c = f(\mathbf{x}) = E[Y | \mathbf{X} = \mathbf{x}] \quad (1.10)$$

f is known as the *regression* function.

2 BAYES CLASSIFIER

2.1 Let (\mathbf{X}, \mathbf{G}) be an input/output dataset, whose relation f is unknown. Also

$$\mathbf{g} \in \mathbf{G} = \{\mathbf{g}_k\}_{k=1}^K \quad (2.1)$$

Let

$$C(\mathbf{g}_k, \mathbf{g}_l) = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases} \quad (2.2)$$

where \mathbf{g}_i are different classes of output data. Thus C is a *correctness* metric.

2.2 Show that

$$\max_{\mathbf{g} \in \mathbf{G}} E[C(\mathbf{G}, f(\mathbf{X}))] = \max_{\mathbf{g} \in \mathbf{G}} p(\mathbf{g} | \mathbf{X} = \mathbf{x}) \quad (2.3)$$

Solution: In the above,

$$\begin{aligned} \max_{\mathbf{g} \in \mathbf{G}} E[C(\mathbf{G}, f(\mathbf{X}))] \\ = \max_{\mathbf{g} \in \mathbf{G}} E_{\mathbf{X}} [E_{\mathbf{G}} \{C(\mathbf{G}, f(\mathbf{x}))\}] \quad (2.4) \end{aligned}$$

*The author is with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India e-mail: gadepall@iith.ac.in. All content in this manual is released under GNU GPL. Free and open source.

$$= \max_{\mathbf{g} \in \mathbf{G}} \sum_{k=1}^K L\{\mathbf{g}_k, \mathbf{g}\} p(\mathbf{g}_k | \mathbf{X} = \mathbf{x}) \quad (2.5)$$

where

$$\sum_{k=1}^K p(\mathbf{g}_k | \mathbf{X} = \mathbf{x}) = 1 \quad (2.6)$$

2.3 Repeat the exercise for the least squares method.

3 EXERCISES

3.1 Explain how (1.10) can be used to obtain the Nearest Neighbour approximation.

3.2 Repeat the exercise for the least squares method.