Statistical Decision Theory

G V V Sharma*

Contents

***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

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

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

$$L\left(Y, \hat{Y}\right) = L(f) \tag{1.2}$$

1.2 Let

$$L(f) = \left(Y - \hat{Y}\right)^2 \tag{1.3}$$

Then

$$E\left[L(f)\right] \tag{1.4}$$

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

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

1.3 Let

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

Using (1.5)

$$\min E\left[L(f)\right) = \min L(f)|\mathbf{X}$$
$$= \min_c E_Y\left\{\left[Y - c\right]^2 |\mathbf{X} = \mathbf{x}\right\} \tag{1.7}$$

Show that

$$L(f)|\mathbf{X} = E_Y\left\{\left[Y - c\right]^2 |\mathbf{X} = \mathbf{x}\right\}$$
$$= -2cE_Y\left\{Y|\mathbf{X} = \mathbf{x}\right\} + E_Y\left\{Y^2|\mathbf{X} = \mathbf{x}\right\} + c^2 \tag{1.8}$$

1.4 $L(f)$ is minimum when

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

Show that this results in

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

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

$$\mathbf{g} \in \mathbf{G} = \left\{\mathbf{g}_k\right\}_{k=1}^K \tag{2.1}$$

Let

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

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

2.2 Show that

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

Solution: In the above,

$$\max_{\mathbf{g}\in\mathbf{G}} E\left[C(\mathbf{G}, f(\mathbf{X}))\right]$$
$$= \max_{\mathbf{g}\in\mathbf{G}} E_{\mathbf{X}}\left[E_{\mathbf{G}}\left\{C(\mathbf{G}, f(\mathbf{x}))\right\}\right] \tag{2.4}$$

*The author is with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India e-mail: gadepall@iith.ac.in.

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

From (2.2), the above expression simplifies to

$$\max_{\mathbf{g} \in \mathbf{G}} E\left[C\left(\mathbf{G}, f\left(\mathbf{X}\right)\right)\right] = \max_{\mathbf{g} \in \mathbf{G}} p\left(\mathbf{g} | \mathbf{X} = \mathbf{x}\right) \tag{2.6}$$

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.