Chapter 1 - Introduction

2. Probability Theory

The Rules of Probability

sum rule $$p(X) = \sum_{Y} p(X, Y)$$

product rule $$p(X, Y) = p(Y|X)p(X).$$

Bayes' Theorem: $$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

2.1. Probability Densities

$$\begin{aligned} p(x) &\geqslant 0 \\ \int_{-\infty}^{\infty} p(x)\,\mathrm{d}x &= 1. \end{aligned}$$

$$p(x) = \int p(x, y)\,\mathrm{d}y$$

2.2. Expectations and Covariances

$$\mathbb{E}[f] = \sum_{x} p(x) f(x)$$

$$\mathbb{E}[f] = \int p(x) f(x)\,\mathrm{d}x.$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2.$$

$$\begin{aligned} \mathrm{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

5. Decision Theory

5.1. Minimising the Misclassification Rate

decision regions - subdivisions of the input space

R1 -> C1; R2 -> C2

decision boundary/surface - border between decision regions

$$\begin{aligned} p(\text{mistake}) &= p(\mathbf{x} \in \mathcal{R}_1, \mathcal{C}_2) + p(\mathbf{x} \in \mathcal{R}_2, \mathcal{C}_1) \\ &= \int_{\mathcal{R}_1} p(\mathbf{x}, \mathcal{C}_2)\,\mathrm{d}\mathbf{x} + \int_{\mathcal{R}_2} p(\mathbf{x}, \mathcal{C}_1)\,\mathrm{d}\mathbf{x}. \end{aligned}$$

$$\begin{aligned} p(\text{correct}) &= \sum_{k=1}^{K} p(\mathbf{x} \in \mathcal{R}_k, \mathcal{C}_k) \\ &= \sum_{k=1}^{K} \int_{\mathcal{R}_k} p(\mathbf{x}, \mathcal{C}_k)\,\mathrm{d}\mathbf{x} \end{aligned}$$

red -> misclassified as C1

green -> misclassified as C1

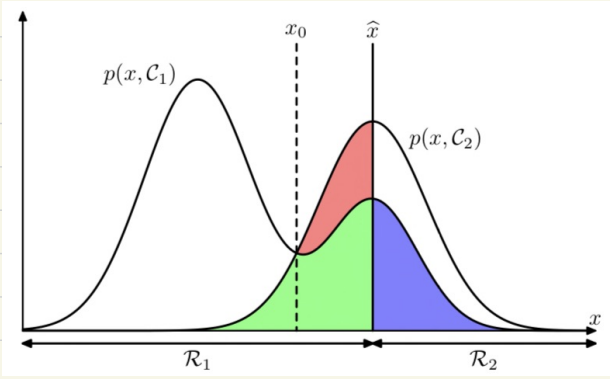
blue -> misclassified as C2

green + blue = const.

red - varies

minimise error=red

=> boundary x0 (red = 0)



5.2. Minimising the Expected Loss

loss/cost function - overall measure of loss due to taking a decision

utility function - negative of the loss

Lkj - loss matrix

$$\mathbb{E}[L] = \sum_k \sum_j \int_{\mathcal{R}_j} L_{kj} p(\mathbf{x}, \mathcal{C}_k)\,\mathrm{d}\mathbf{x}.$$

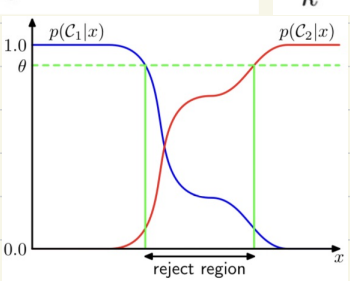
$$\sum_k L_{kj} p(\mathcal{C}_k|\mathbf{x})$$

5.3. The Reject Option

avoid making decisions

K - number of classes

1 > θ > 1/K



5.4. Inference and Decision

discrimination function - inference + decision

outlier/novelty detection - low accuracy predictions

generative models - model input as well as output

discriminative models - model posterior probabilities p(Clx)