CS391L Assignment 3 Problem Set

The problems are to be solved on your own without consulting any source of answers. The best submission format would be some kind of pdf uploaded as a file. The numbers in brackets refer to equations from Bishop's PRML text.

1.1 Consider the sum-of-squares error function given by (1.2) in which the function $y(x, \mathbf{w})$ is given by the polynomial (1.1). Show that the coefficients $\mathbf{w} = \{w_i\}$ that minimize this error function are given by the solution to the following set of linear equations

$$\sum_{j=0}^{M} A_{ij} w_j = T_i \tag{1.122}$$

where

$$A_{ij} = \sum_{n=1}^{N} (x_n)^{i+j}, \qquad T_i = \sum_{n=1}^{N} (x_n)^i t_n. \tag{1.123}$$

Here a suffix i or j denotes the index of a component, whereas $(x)^i$ denotes x raised to the power of i.

1.2 ($\star$) Write down the set of coupled linear equations, analogous to (1.122), satisfied by the coefficients w_i which minimize the regularized sum-of-squares error function given by (1.4).

1.3 ($\star\star$) Suppose that we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r) = 0.2$, $p(b) = 0.2$, $p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

1.39 ($\star\star\star$) Consider two binary variables x and y having the joint distribution given in Table 1.3.

Evaluate the following quantities

(a) $\mathrm{H}[x]$	**(c)** $\mathrm{H}[y\|x]$	**(e)** $\mathrm{H}[x, y]$
(b) $\mathrm{H}[y]$	**(d)** $\mathrm{H}[x\|y]$	**(f)** $\mathrm{I}[x, y]$.

Draw a diagram to show the relationship between these various quantities.

1.40 ($\star$) By applying Jensen's inequality (1.115) with $f(x) = \ln x$, show that the arithmetic mean of a set of real numbers is never less than their geometrical mean.

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \leqslant \sum_{i=1}^{M} \lambda_i f(x_i) \tag{1.115}$$

where $\lambda_i \geqslant 0$ and $\sum_i \lambda_i = 1$, for any set of points $\{x_i\}$. The result (1.115) is known as *Jensen's inequality*.

The arithmetic and geometric means are defined as

$$\bar{x}_{\mathrm{A}} = \frac{1}{K}\sum_{k}^{K} x_k \quad \text{and} \quad \bar{x}_{\mathrm{G}} = \left(\prod_{k}^{K} x_k\right)^{1/K},$$

1.41 Using the sum and product rules of probability, show that the mutual information $I(\mathbf{x}, \mathbf{y})$ satisfies the relation (1.121).

Decision Tree question

Example	crust size	shape	filling size	Class
e1	big	circle	small	**pos**
e2	small	circle	small	**pos**
e3	big	square	small	**neg**
e4	big	triangle	small	**neg**
e5	big	square	big	**pos**
e6	small	square	small	**neg**
e7	small	square	big	**pos**
e8	big	circle	big	**pos**

Here is the entropy of the training set where only class labels are known:

$$\begin{aligned} H(T) &= -p_{\text{pos}} \log_2 p_{\text{pos}} - p_{\text{neg}} \log_2 p_{\text{neg}} \\ &= -(5/8)\log(5/8) - (3/8)\log(3/8) = 0.954 \end{aligned}$$

In building the tree, what attribute should be selected for the first node?
Show all calculations

7.4 Show that the value ρ of the margin for the maximum-margin hyperplane is given by

$$\frac{1}{\rho^2} = \sum_{n=1}^{N} a_n \tag{7.123}$$

where $\{a_n\}$ are given by maximizing (7.10) subject to the constraints (7.11) and (7.12).

11.11 Show that the Gibbs sampling algorithm, discussed in Section 11.3, satisfies detailed balance as defined by (11.40).