

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

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

where

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

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

Prove:

$$E(w) = \frac{1}{2} \sum_{n=1}^N (y(x_n, \mathbf{w}) - t_n)^2$$

$$\frac{\partial E}{\partial w_i} = \sum_{n=1}^N (y(x, \mathbf{w}) - t_n) \cdot x_n^i = 0$$

so,

$$\begin{aligned} \sum_{n=1}^N (y(x, \mathbf{w}) - t_n) \cdot x_n^i &= 0 \\ \sum_{n=1}^N y(x, \mathbf{w}) \cdot x_n^i - \sum_{n=1}^N t_n \cdot x_n^i &= 0 \\ \sum_{n=1}^N y(x, \mathbf{w}) \cdot x_n^i &= \sum_{n=1}^N t_n \cdot x_n^i \\ \sum_{n=1}^N \left( \sum_{j=0}^M w_j x_n^j \right) \cdot x_n^i &= \sum_{n=1}^N t_n \cdot x_n^i \\ \sum_{n=1}^N \sum_{j=0}^M w_j x_n^{j+i} &= \sum_{n=1}^N t_n \cdot x_n^i \\ \sum_{j=0}^M w_j \sum_{n=1}^N x_n^{j+i} &= \sum_{n=1}^N x_n^i \cdot t_n \end{aligned}$$

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

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

**1.2 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)**

*Sol:*

$$E(w) = \frac{1}{2} \sum_{n=1}^N (y(x_n, \mathbf{w}) - t_n)^2 + \frac{1}{\lambda} \|\mathbf{w}\|^2$$

$$\|\mathbf{w}\|^2 = w_0^2 + w_1^2 + \dots + w_M^2$$

$$\frac{\partial E}{\partial w_i} = \sum_{n=1}^N (y(x, \mathbf{w}) - t_n) \cdot x_n^i + \lambda w_i = 0$$

$$\sum_{n=1}^N (y(x, \mathbf{w}) - t_n) \cdot x_n^i + \lambda w_i = 0$$

$$\sum_{n=1}^N y(x, \mathbf{w}) \cdot x_n^i - \sum_{n=1}^N t_n \cdot x_n^i + \lambda w_i = 0$$

$$\sum_{n=1}^N y(x, \mathbf{w}) \cdot x_n^i + \lambda w_i = \sum_{n=1}^N t_n \cdot x_n^i$$

$$\sum_{n=1}^N \left( \sum_{j=0}^M w_j x_n^j \right) \cdot x_n^i + \lambda w_i = \sum_{n=1}^N t_n \cdot x_n^i$$

$$\sum_{n=1}^N \sum_{j=0}^M w_j x_n^{j+i} + \lambda w_i = \sum_{n=1}^N t_n \cdot x_n^i$$

$$\sum_{j=0}^M w_j \left( \sum_{n=1}^N x_n^{j+i} + \lambda I_{ij} \right) = \sum_{n=1}^N x_n^i \cdot t_n$$

$I_{ij}$  denotes the identity matrix.

$$\text{So, } \tilde{A}_{ij} = \sum_{n=1}^N (x_n)^{i+j} + \lambda I_{ij}, \quad T_i = \sum_{n=1}^N (x_n)^i t_n$$

$$\therefore \sum_{j=0}^M \tilde{A}_{ij} w_j = T_i$$

**1.3 Suppose we have three coloured boxes  $r$  (red),  $b$  (blue), and  $g$  (green), box  $r$  contains 3 apples, 4 oranges, and 3 lime, box  $b$  contains 1 apples, 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**

*Sol:*

$$\begin{aligned}
P(\text{apple}) &= P(\text{apple}|r) \cdot P(r) + P(\text{apple}|b) \cdot P(b) + P(\text{apple}|g) \cdot P(g) \\
&= \frac{3}{10} * 0.2 + \frac{1}{2} * 0.2 + \frac{3}{10} * 0.6 \\
&= 0.34
\end{aligned}$$

$$\begin{aligned}
P(g|\text{orange}) &= \frac{P(\text{orange}, g)}{P(\text{orange})} \\
&= \frac{P(\text{orange}, g)}{P(\text{orange}|r) \cdot P(r) + P(\text{orange}|b) \cdot P(b) + P(\text{orange}|g) \cdot P(g)} \\
&= \frac{0.3 * 0.6}{0.4 * 0.2 + 0.5 * 0.2 + 0.3 * 0.6} \\
&= \frac{1}{2} = 0.5
\end{aligned}$$

**1.6 Show that if two variables  $x$  and  $y$  are independent, then their covariance is zero**

*Prove:*

$$\begin{aligned}
cov(x, y) &= E[(x - E[x]) \cdot (y - E[y])] \\
&= E[xy - yE[x] - xE[y] + E[x]E[y]] \\
&= E[xy] - E[y]E[x] - E[x]E[y] + E[x]E[y] \\
&= E[xy] - E[x]E[y]
\end{aligned}$$

$x, y$  is independent, so  $E[xy] = E[x]E[y]$

so,  $cov(x, y) = 0$

**1.10 Suppose that the two variables  $x$  and  $z$  are statistically independent. Show that the mean and variance of their sum satisfies**

$$\begin{aligned}
E[x + z] &= E[x] + E[z] \\
var[x + z] &= var[x] + var[z]
\end{aligned}$$

*Prove:*

i). For any two variables  $x, z$ , they satisfy the condition  $E[x + z] = E[x] + E[z]$

ii).

$$\begin{aligned}
&\because x, z \text{ is independent} \\
&\therefore cov(x, z) = 0 \\
&\therefore var[x + z] = var[x] + var[z] + cov(x, z) \\
&\quad = var[x] + var[z]
\end{aligned}$$

**1.17 The gamma function is defined by**

$$\Gamma(x) \equiv \int_0^{\infty} u^{x-1} e^{-u} du$$

Using integration by parts, prove the relation  $\Gamma(x + 1) = x\Gamma(x)$ . Show also that  $\Gamma(1) = 1$  and hence that  $\Gamma(x + 1) = x!$  when  $x$  is an integer

*Prove:*

$$\begin{aligned}
\Gamma(x) &= \int_0^{\infty} u^{x-1} e^{-u} du \\
&= \frac{1}{x} \int_0^{\infty} x u^{x-1} e^{-u} du \\
&= \frac{1}{x} \int_0^{\infty} e^{-u} du^x \\
&= \frac{1}{x} e^{-u} u^x \Big|_0^{\infty} - \frac{1}{x} \int_0^{\infty} u^x d e^{-u} \\
&= 0 - \frac{1}{x} \int_0^{\infty} u^x d e^{-u} \\
&= -\frac{1}{x} \int_0^{\infty} u^x d e^{-u}
\end{aligned}$$

$$\begin{aligned}
\Gamma(x+1) &= \int_0^{\infty} u^x e^{-u} du \\
&= - \int_0^{\infty} u^x d e^{-u} \\
&= x \Gamma(x)
\end{aligned}$$

So,  $\Gamma(x+1) = x\Gamma(x)$

$$\Gamma(1) = \int_0^{\infty} e^{-u} du = -e^{-u} \Big|_0^{\infty} = -(0 - 1) = 1$$

$$\Gamma(x+1) = x\Gamma(x) = x(x-1)\Gamma(x-1) = \dots = x(x-1) \dots 1\Gamma(1) = x!$$

**1.25 Consider the generalization of squared loss function (1.87) for a single target variable  $t$  to the case of multiple target variables described by the vector  $\mathbf{t}$  given by**

$$E[L(\mathbf{t}, \mathbf{y}(\mathbf{x}))] = \int \int \|\mathbf{y}(\mathbf{x}) - \mathbf{t}\|^2 p(\mathbf{x}, \mathbf{t}) d\mathbf{x} d\mathbf{t}$$

**Using the calculus of variations, show that the function  $y(x)$  for which this expected loss minimized is given by  $y(x) = E_t[t|x]$ . Show that this result reduces to (1.89) for the case of a single target variable  $t$**

*Prove:*

Using the calculus of variations to give

$$\frac{\delta E[L]}{\delta \mathbf{y}(\mathbf{x})} = \int 2(\mathbf{y}(\mathbf{x}) - \mathbf{t}) p(\mathbf{t}, \mathbf{x}) d\mathbf{t} = 0$$

Solving for  $\mathbf{y}(\mathbf{x})$ , and using the sum and product rules and probability, we obtain

$$\mathbf{y}(\mathbf{x}) = \frac{\int \mathbf{t} p(\mathbf{t}, \mathbf{x}) d\mathbf{t}}{\int p(\mathbf{t}, \mathbf{x}) d\mathbf{t}} = \int \mathbf{t} p(\mathbf{t}|\mathbf{x}) d\mathbf{t}$$

which is the conditional average if  $\mathbf{t}$  conditioned on  $\mathbf{x}$ . For the case of a scalar target variable we have

$$y(\mathbf{x}) = \int t p(t|\mathbf{x}) dt$$

**1.30 Evaluate the Kullback-Leibler divergence between two Gaussians  $p(x) = N(x|\mu, \sigma^2)$  and  $q(x) = N(x|m, s^2)$ .**

*Sol:*

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$q(x) = \frac{1}{\sqrt{2\pi}s} e^{-\frac{(x-m)^2}{2s^2}}$$

$$\frac{p(x)}{q(x)} = \frac{s}{\sigma} e^{\frac{(x-m)^2}{s^2} - \frac{(x-\mu)^2}{\sigma^2}}$$

$$\frac{(x-m)^2}{s^2} - \frac{(x-\mu)^2}{\sigma^2} = \left(\frac{1}{s^2} + \frac{1}{\sigma^2}\right)x^2 + \left(\frac{2\mu}{\sigma^2} - \frac{2m}{s^2}\right)x + \left(\frac{m^2}{s^2} + \frac{\mu^2}{\sigma^2}\right)$$

$$\begin{aligned} KL(p|q) &= \int p(x) \ln p(x) dx - \int p(x) \ln q(x) dx \\ &= \int p(x) \ln \left( \frac{p(x)}{q(x)} \right) dx \\ &= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln \left( \frac{s}{\sigma} e^{\frac{(x-m)^2}{s^2} - \frac{(x-\mu)^2}{\sigma^2}} \right) dx \\ &= \int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln \left( \frac{s}{\sigma} \right) + \int N(x|\mu, \sigma^2) \left[ \left( \frac{1}{s^2} + \frac{1}{\sigma^2} \right) x^2 + \left( \frac{2\mu}{\sigma^2} - \frac{2m}{s^2} \right) x + \left( \frac{m^2}{s^2} + \frac{\mu^2}{\sigma^2} \right) \right] dx \\ &= \ln \left( \frac{s}{\sigma} \right) + \left( \frac{1}{s^2} + \frac{1}{\sigma^2} \right) \int N(x|\mu, \sigma^2) x^2 dx + \left( \frac{2\mu}{\sigma^2} - \frac{2m}{s^2} \right) \int N(x|\mu, \sigma^2) x dx \\ &\quad + \left( \frac{m^2}{s^2} + \frac{\mu^2}{\sigma^2} \right) \int N(x|\mu, \sigma^2) dx \\ &= \ln \left( \frac{s}{\sigma} \right) + \left( \frac{1}{s^2} + \frac{1}{\sigma^2} \right) (\mu^2 + \sigma^2) + \left( \frac{2\mu}{\sigma^2} - \frac{2m}{s^2} \right) (\mu) + \left( \frac{m^2}{s^2} + \frac{\mu^2}{\sigma^2} \right) \end{aligned}$$

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

*Sol:*

Assume real number  $x_1, x_2, \dots, x_n$

The arithmetic mean of these real number is :

$$\bar{x}_A = \frac{1}{n} \sum_{i=1}^n x_i$$

$$f(\bar{x}_A) = \ln \left( \frac{1}{n} \sum_{i=1}^n x_i \right)$$

The geometrical mean of these real numbers is:

$$\bar{x}_G = \left( \prod_{i=1}^n x_i \right)^{\frac{1}{n}}$$

$$f(\bar{x}_G) = \ln\left(\left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}}\right) = \frac{1}{n} \sum_{i=1}^n \ln(x_i)$$

from

$$f\left(\sum_i^M \lambda_i x_i\right) \leq \sum_i^M \lambda_i f(x_i)$$

let  $\lambda_i = \frac{1}{n}$ , so  $f(\bar{x}_A) \leq f(\bar{x}_G)$

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

$$I[x, y] = H[x] - H[x|y]$$

$$H[x] = - \int p(x) \ln(p(x)) dx$$

$$\begin{aligned} H[x] - H[x|y] &= - \int p(x) \ln[p(x)] dx + \int \int p(x, y) \ln[p(x|y)] dx dy \\ &= \int \int p(x, y) (\ln[p(x, y)] - \ln[p(y)]) dx dy - \int p(x) \ln(p(x)) dx \\ &= \int \int p(x, y) \ln[p(x, y)] dx dy - \int \int p(x, y) \ln[p(y)] dx dy - \int p(x) \ln(p(x)) dx \\ &= \int \int p(x, y) \ln[p(x, y)] dy dx - \int p(y) \ln[p(y)] dy - \int \int p(x, y) \ln[p(x)] dy dx \\ &= \int \int p(x, y) \ln\left[\frac{p(x, y)}{p(x)}\right] dy dx - \int p(y) \ln[p(y)] dy \\ &= \int \int p(x, y) \ln[p(y|x)] dy dx - \int p(y) \ln[p(y)] dy \\ &= H[y] - H[y|x] \end{aligned}$$