AML

Exercise 1.1

(a) The input space X → Medical history and some symptoms
The output space Y → Is the patient sick (Yes or No)
The target function f → Discriminant function, which is a mapping of patient features and a label of yes or no.

(b) The input space X → Handwritten digit picture
The output space Y → Recognized digit
The target function f → Discriminant function, which is a mapping of handwritten digit and digit.

(c) The input space X → Email content
The output space Y → An email is spam or not (Yes or No)
The target function f → Discriminant function, which is a mapping of email content and a label of yes or no.

(d) The input space X → Electric load, price, temperature, and day of the week
The output space Y → Electricity
The target function f → Predictive function.

(e) The input space X → The data you have
The output space Y → Results of your problem
The target function f → Empirical solution.

Exercise 1.2

(a) Money, personal information, gambling etc.
(b) Some information not related to spam, such as work, homework, schedule etc.
(c) The parameter $w_i (i = 1,2..n)$ and bias b in perceptron model could affect borderline messages.

Exercise 1.3

The equation (1) is perceptron model and we can convert that to equation (2) by setting $w_0 = b$ and $x_0 = 1$,

$$\mathrm{h(x)} = sign\left(\left(\sum_{i=1}^{d} w_i x_i\right) + b\right) \quad (1)$$

$$\mathrm{h(x)} = sign\left(\left(\sum_{i=0}^{d} w_i x_i\right)\right) = sign(w^T(t)x(t)) \quad (2)$$

According to equation 1.3 in book page 7, we can see that if $\mathrm{y(t)} \neq sign(w^T(t)x(t))$, then $w(t+1) = w(t) + y(t)x(t)$

(a) $\because y(t) \neq sign(w^T(t)x(t))$
$\therefore$ if $sign\big(w^T(t)x(t)\big) > 0$ then $\mathrm{y(t)} = -1$, if $sign\big(w^T(t)x(t)\big) < 0$ then $y(t) = 1$
$\therefore\ y(t)w^T(t)x(t) < 0$

(b) We can make the following derivation:

$$y(t)w^T(t+1)x(t) = y(t)\big(w(t) + y(t)x(t)\big)x(t) = y(t)w^T(t)x(t) + y^2(t)x^T(t)x(t)$$

$\because$ The first component of x(t) is 1

$\therefore$ $y^2(t)x^T(t)x(t) > 0$

$\therefore$ $y(t)w^T(t+1)x(t) > y(t)w^T(t)x(t)$

(c) We can see that if the classification is wrong of a sample, $y(t)w^T(t)x(t) < 0$ by problem (a), but according to equation 1.3 in book page 7, if $y(t) \neq sign(w^T(t)x(t))$, then $w(t+1) = w(t) + y(t)x(t)$, so, if our data is assortable, w could be calculated and make $yw^Tx > 0$ for all x.

Exercise 1.4

Please refer to the attachment (Jupyter notebook -- **AML-Exercise 1.4.ipynb**) for detailed code.

Exercise 1.5

(a) Learning approach
(b) Design approach
(c) Learning approach
(d) Design approach
(e) Learning approach

Exercise 1.6

(a) Unsupervised learning
(b) Win or lose is certain of the game, so, it is a supervised learning, and the learning process is reinforcement learning.
(c) It is a kind of classification problem, so it is supervised learning.
(d) I is a supervised learning, and the learning process is reinforcement learning.
(e) It is a kind of regression prediction problem, so it is supervised learning.

Exercise 1.10

Please refer to the attachment (Jupyter notebook -- **AML-Exercise 1.10.ipynb**) for detailed code.

PRML

Exercise 1.1

$$y(x, w) = \sum_{j}^{M} w_j x^j$$

$$E(x) = \frac{1}{2}\sum_{n=1}^{N}(y(x_n, w) - t_n)^2 = \frac{1}{2}\sum_{n=1}^{N}\left(\sum_{j=0}^{M} w_j x_n^j - t_n\right)^2$$

$$\frac{\partial}{\partial w_i}E(w) = \sum_{n=1}^{N}\left(\sum_{j=0}^{M} w_j x_n^j - t_n\right)x_n^i = 0$$

$$\Downarrow$$

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

$$\Downarrow$$

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

Exercise 1.5

$$E[(f(x) - E[f(x)])^2]$$

$$= \int [f(x)^2 - 2f(x)E[f(x)] + E[f(x)]^2]p(x)dx$$

$$= \int f(x)^2 p(x)dx$$

$$- 2E[f(x)]\int f(x)p(x)dx + E[f(x)^2]\int p(x)dx = E[f(x)^2] - E[f(x)]^2$$

Exercise 1.6

$$E_{x,y}[xy] - E[x]E[y] = \sum_x\sum_y xyp(xy) - \sum_x xp(x)\sum_y yp(y) \quad (3)$$

∵ x and y are independent of each other

∴ $p(xy) = p(x)p(y)$

∴ equation (3) equals to 0

Exercise 1.9

(1) Univariate:

We can get equation (4) by (1.46):

$$\frac{d}{dx}N(x|\mu,\sigma^2) = -N(x|\mu,\sigma^2)\frac{x-\mu}{\sigma^2} \quad (4)$$

Then, we set (4) to zero, and we can get $x = \mu$

(2) Multivariate:

We can get equation (5) by (1.52) and (C.19) (C.20):

$$\frac{\partial}{\partial x}N(x|\mu,\Sigma) = -\frac{1}{2}N(x|\mu,\Sigma)\nabla_x\{(x-\mu)^T\Sigma^{-1}(x-\mu)\}$$

$$= -N(x|\mu,\Sigma)\,\Sigma^{-1}(x-\mu)) \quad (5)$$

Then, we set (5) to zero, and we can get $x = \mu$

Exercise 1.11

(1) Proof:

$$\frac{\partial\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2-\frac{N}{2}ln\sigma^2-\frac{N}{2}\ln(2\pi)\right\}}{\partial \mathrm{u}}-\frac{1}{\sigma^2}\sum_{n=1}^{N}(\mu-x_n)=0$$

$$\Downarrow$$

$$\mu=\frac{1}{N}\sum_{n=1}^{N}x_n$$

(2) Proof:

$$\frac{\partial\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2-\frac{N}{2}ln\sigma^2-\frac{N}{2}\ln(2\pi)\right\}}{\partial \sigma^2}-\frac{1}{\sigma^2}\sum_{n=1}^{N}(\mu-x_n)=0$$

$$\Downarrow$$

$$\mathrm{t}=\frac{1}{N}\sum_{n=1}^{N}(x_n-\mu)^2$$

Exercise 1.12

Proof:

According to (1.50) and assuming m = n, we can get $E[x_n^2]=\mu^2+\sigma^2$.

However, according to (1.49), if m≠n, we can get $E[x_n,x_m]=E[x_n]E[x_m]=\mu^2$, due to the sample data x_n and x_m are independent.

Equation (1.130) can be obtained by combining above inferences.

Considering equation (1.49) we can get equation (6) below:

$$E[\mu_{ML}]=\frac{1}{N}\sum_{n=1}^{N}E[x_n]=\mu \quad (6)$$

Considering equation (1.55), (1.56) and equation (1.130), we can make the following derivation:

$$E[\sigma_{ML}^2]=E\left[\frac{1}{N}\sum_{n=1}^{N}\left(x_n-\frac{1}{N}\sum_{m=1}^{N}x_m\right)^2\right]$$

$$=\frac{1}{N}\sum_{n=1}^{N}E[x_n^2-\frac{2}{N}x_n\sum_{m=1}^{N}x_m+\frac{1}{N^2}\sum_{m=1}^{N}\sum_{l=1}^{N}x_m x_l]$$

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

$$=\left(\frac{N-1}{N}\right)\sigma^2 \qquad (7)$$

In summary, (1.57) and (1.58) have been proved.