SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Homework #1

Course: *Machine Learning (CS405)* – Professor: *Qi Hao*
Due date: *11:59pm, September 23th, 2020*

NAME: 车凯威
NUMBER: 12032207

Question 1

Consider the polynomial function

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x + ... + w_M x^M = \sum_{j=0}^{M} w_j x^j$$

Calculate the coefficients $\mathbf{w} = \{w_i\}$ that minimize its sum-of-squares error function. Here a suffix i or j denotes the index of a component, whereas $(x)^i$ denotes x raised to the power of i.

Question 2

Suppose that we have three colored 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?

Question 3

Given two statistically independent variables x and z, show that the mean and variance of their sum satisfies

$$\mathbb{E}[x + z] = \mathbb{E}[x] + \mathbb{E}[z]$$

$$\text{var}[x + z] = \text{var}[x] + \text{var}[z]$$

Question 1:

Assume that: M means the order of the polynomial
N means N samples
t_n means the real value of the input of $x^{(n)}$
$y(x, w)$ means the predict value of the input of $x^{(n)}$

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

Gradient Descent: $w_j := w_j - \alpha \frac{\partial}{\partial w_j} E(w)$

$$\frac{\partial}{\partial w_j} E(w) = \sum_{n=1}^{N} (y(x_n, w) - t_n) \cdot x_j^{(n)}$$

Therefore, the w_j that minimize its sum-of-squares-error function is:

$$w_j := w_j - \alpha \cdot \sum_{n=1}^{N} (y(x_n, w) - t_n) \cdot x_j^{(n)}, \quad j \in M$$

Question 2:

Assume event F means the probability of selecting fruit from box.

a is apple, o is orange, l is limes

Assume event B means the probability of selecting box

1. $p(F=a) = \sum_B p(F=a \mid B)$

$$= p(F=a \mid B=r)p(B=r) + p(F=a \mid B=b)p(B=b)$$
$$+ p(F=a \mid B=g)p(B=g)$$
$$= \frac{2}{10} \cdot \frac{3}{10} + \frac{2}{10} \cdot \frac{1}{2} + \frac{6}{10} \cdot \frac{3}{10}$$
$$= \frac{34}{100} = 0.34$$

2. According to Bayes' theorem:

$$p(B=g \mid F=o) = \frac{p(F=o \mid B=g)\,p(B=g)}{p(F=o)}$$

$$p(F=o \mid B=g) = \frac{3}{10} = 0.3$$

$$p(B=g) = 0.6$$

$$p(F=o) = \sum_B p(F=o \mid B)$$
$$= p(F=o \mid B=r)p(B=r) + p(F=o \mid B=b)p(B=b)$$
$$+ p(F=o \mid B=g)p(B=g)$$
$$= \frac{4}{10} \times \frac{2}{10} + \frac{1}{2} \times \frac{2}{10} + \frac{3}{10} \times \frac{6}{10}$$
$$= \frac{36}{100} = 0.36$$

$$p(B=g \mid F=o) = \frac{p(F=o \mid B=g)\,p(B=g)}{p(F=o)}$$

$$= \frac{0.3 \times 0.6}{0.36} = \frac{1}{2}$$

Question 3:

$$E(x+z) = E(x) + E(z)$$

Discrete:

$$E[x] = \sum_{i=1}^{M} x_i \, p(x_i)$$

$$E[z] = \sum_{j=1}^{N} z_j \, P(z_j)$$

$$E[x+z] = \sum_{i=1}^{M} \sum_{j=1}^{N} (x_i + z_j) P(x_i z_j)$$

$$= \sum_{i=1}^{M} \sum_{j=1}^{N} (x_i + z_j) P(x_i) P(z_j)$$

$$= \sum_{i=1}^{M} x_i P(x_i) + \sum_{j=1}^{N} z_j P(z_j)$$

$$= E[x] + E[z]$$

Continuous:

$$E[x] = \int x f(x) \, dx$$

$$E[z] = \int z f(z) \, dz$$

$$E[x+z] = \iint (x+z) f(xz) \, dx \, dz$$

$$= \iint (x+z) f(x) f(z) \, dx \, dz$$

$$= \iint x f(x) f(z) \, dx \, dz + \iint z f(x) f(z) \, dx \, dz$$

$$= \int x f(x) \, dx + \int z f(z) \, dz$$

$$= E[x] + E[z]$$

$$var[x+z] = var[x] + var[z]$$

$$var[x] = E(x^2) - (Ex)^2$$

$$var[z] = E(z^2) - (Ez)^2$$

$$\begin{aligned}
var[x+z] &= E((x+z)^2) - (E(x+z))^2 \\
&= E(x^2 + 2xz + z^2) - \left[(Ex)^2 + 2ExEz + (Ez)^2\right] \\
&= Ex^2 + 2ExEz + Ez^2 - (Ex)^2 - 2ExEz - (Ez)^2 \\
&= Ex^2 - (Ex)^2 + Ez^2 - (Ez)^2 \\
&= var[x] + var[z]
\end{aligned}$$

Question 4

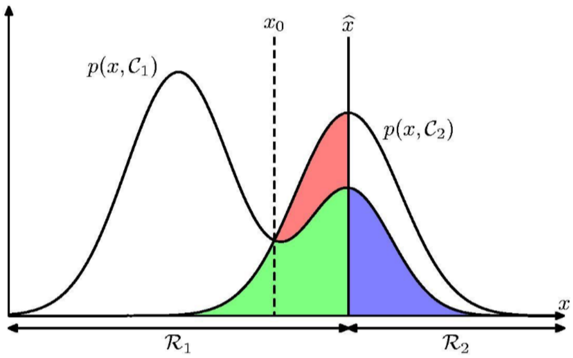
In probability theory and statistics, the Poisson distribution, is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate and independently of the time since the last event. If X is Poisson distributed, i.e. $X \sim Possion(\lambda)$, its probability mass function takes the following form:

$$P(X|\lambda) = \frac{\lambda^X e^{-\lambda}}{X!}$$

It can be shown that if $\mathbb{E}(X) = \lambda$. Assume now we have n data points from $Possion(\lambda) : \mathcal{D} = \{X_1, X_2, ..., X_n\}$. Show that the sample mean $\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the maximum likelihood estimate(MLE) of λ. If X is exponential distribution and its distribution density function is $f(x) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}}$ for $x > 0$ and $f(x) = 0$ for $x \leq 0$. Show that the sample mean $\widehat{\lambda}\frac{1}{n}\sum_{i=1}^{n} X_i$ is the maximum likelihood estimate(MLE) of λ.

Question 5

(*a*) Write down the probability of classifying correctly $p(correct)$ and the probability of misclassification $p(mistake)$ according to the following chart.



(*b*) For multiple target variables described by vector $\mathbf{t}$, the expected squared loss function is given by

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

Show that the function $\mathbf{y}(\mathbf{x})$ for which this expected loss is minimized given by $\mathbf{y}(\mathbf{x}) = \mathbb{E}_{\mathbf{t}}[\mathbf{t}|\mathbf{x}]$.

Hints. For a single target variable t, the loss is given by

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

The result is as follows

$$y(\mathbf{x}) = \frac{\int t p(\mathbf{x}, t)\mathrm{d}t}{p(\mathbf{x})} = \int t p(t|\mathbf{x})\mathrm{d}t = \mathbb{E}_t[t|\mathbf{x}]$$

Question 4:

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

$$= \frac{1}{x_1! \cdot x_2! \cdots x_n!} \cdot \lambda^{\sum_{i=1}^{n} x_i} e^{-\lambda n}$$

$$\ln L(\lambda) = \ln \frac{1}{x_1! \cdots x_n!} + \sum_{i=1}^{n} x_i \ln \lambda - \lambda n$$

$$\frac{d \ln L(\lambda)}{d\lambda} = \sum_{i=1}^{n} x_i \frac{1}{\lambda} - n = 0$$

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Question 5:

(a) $$p(\text{correct}) = \sum_{k=1}^{K} p(x \in R_k, C_k)$$

$$= \sum_{k=1}^{K} \int_{R_k} p(x, C_k)\, dx$$

$$p(\text{mistake}) = p(x \in R_1, C_2) + p(x \in R_2, C_1)$$

$$= \int_{R_1} p(x, C_2)\, dx + \int_{R_2} p(x, C_2)\, dx$$

(b)

$$E[L] = \iint (y(x) - t)^2 p(x,t)\,dx\,dt$$

$$E[L] = \iint \left[y^2(x) - 2y(x)t + t^2 \right] p(x,t)\,dx\,dt$$

$$= y^2(x) \iint p(x,t)\,dx\,dt -$$

$$2y(x) \iint t\,p(x,t)\,dx\,dt +$$

$$\iint t^2 p(x,t)\,dx\,dt$$

$$= y^2(x) - 2y(x) \int t\,p(t|x)\,dt$$

$$\frac{\partial E[L]}{\partial y(x)} = 2y(x) - 2\int t\,p(t|x)\,dt = 0$$

$$y(x) = \int t\,p(t|x)\,dt = E_t(t|x)$$

Because multiple target variables vector t is defined by $t[i]$, therefore vector $y(x) = E_t[t|x]$

Question 6

(*a*) We defined the entropy based on a discrete random variable $\mathbf{X}$ as

$$\mathbf{H}[\mathbf{X}] = -\sum_i p(x_i) \ln p(x_i)$$ the probability.

Now consider the case that $\mathbf{X}$ is a continuous random variable with the probability density function $p(x)$. The entropy is defined as

$$\mathbf{H}[\mathbf{X}] = -\int p(x) \ln p(x) dx$$

Assume that $\mathbf{X}$ follows Gaussian distribution with the mean μ and variance σ, i.e.

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

Please derive its entropy $\mathbf{H}[\mathbf{X}]$.

(*b*) Write down the mutual information $\mathbf{I}(\mathbf{y}|\mathbf{x})$. Then show the following equation

$$\mathbf{I}[\mathbf{x}, \mathbf{y}] = \mathbf{H}[\mathbf{x}] - \mathbf{H}[\mathbf{x}|\mathbf{y}] = \mathbf{H}[\mathbf{y}] - \mathbf{H}[\mathbf{y}|\mathbf{x}]$$

Question 6:

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

$$= \int \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{(x-\mu)^2}{2\delta^2}} \cdot \left[\ln\sqrt{2\pi}\delta + \frac{(x-\mu)^2}{2\delta^2}\right] dx$$

$$= \frac{\ln\sqrt{2\pi}\delta}{\sqrt{2\pi}\delta} \int e^{-\frac{(x-\mu)^2}{2\delta^2}} dx + \frac{1}{\sqrt{2\pi}\delta} \int e^{-\frac{(x-\mu)^2}{2\delta^2}} \cdot \frac{(x-\mu)^2}{2\delta^2} dx$$

$t = \frac{x-\mu}{\delta}$,

$$= \frac{\sqrt{2}}{2} \frac{\ln 2\pi}{\pi} + \frac{1}{\sqrt{2\pi}} \underbrace{\int e^{-\frac{t^2}{2}} \frac{t^2}{2} dt}$$

$$e^{-\frac{t^2}{2}} \cdot t^2 + \int e^{-\frac{t^2}{2}} dt = e^{-\frac{t^2}{2}} \cdot t^2 + 1$$

$$= \frac{\sqrt{2}}{2} \cdot \frac{\ln 2\pi}{\pi} + \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\delta^2}} \cdot \frac{(x-\mu)^2}{\delta^2} + \frac{1}{\sqrt{2\pi}}$$

(b) $I(x,y) = \sum_x \sum_y P(x,y) \ln \frac{P(x,y)}{P(x)P(y)}$

$H(x) = -\sum_x P(x)\ln P(x)$

$H(y) = -\sum_y P(y)\ln P(y)$

$$I(x,y) = \sum_x \sum_y P(x,y) \ln \frac{P(x,y)}{P(y)} - \sum_x P(x)\ln P(x)$$

$$= \sum_x \sum_y P(y)P(x|y)\ln P(x|y) + H(x)$$

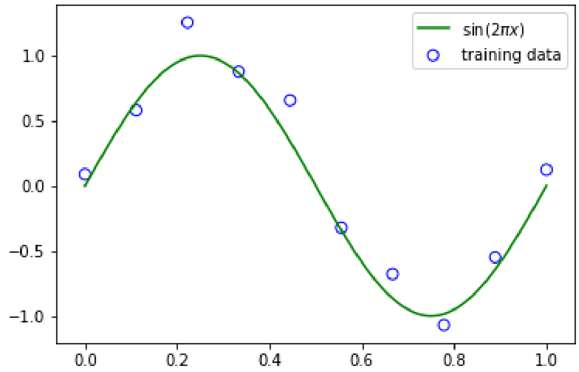
$$= -H(x|y) + H(x) = H(x) - H(x|y)$$

the same : $I(x,y) = H(y) - H(y|x)$

Program Question

You should download the `HW1_programQuestion.ipynb` file first.

(*a*) Plot the graph with given code, the result should be same as this.



(*b*) On the basis of the results, you should try 0^{th} order polynomial, 1^{st} order polynomial, 3^{rd} order polynomial and some other order polynomial, show the results include fitting and over-fitting.

(*c*) Plot the graph of the root-mean-square error.

(*d*) Plot the graph of the predictive distribution resulting from a Bayesian treatment of polynomial curve fitting using an M=9 polynomial, with the fixed parameters $\alpha = 5 \times 10^{-3}$ and $\beta = 11.1$(corresponding to the known noise variance).

(*e*) Change the *sample_size* to 2, 3 or 10 times than before, explain the change of M.

Hints. You should install `matplotlib.pyplot`, and read classes `PolynomialFeature`, `LinearRegression`, and `BayesianRegression` in the file.