Homework 1 - Written Answer Key

Question 1:

(Question)

"Explain whether each scenario is a classification or regression problem, and provide N and d (d is the number of features).

(a) We collect a set of data on the top 500 firms in the US. For each firm we record profit, number of employees, industry and the CEO salary. We are interested in understanding which factories affect CEO salary.

(b) We are considering launching a new product and wish to know whether it will be a *success* or *failure*. We collect data on 20 similar products that were previously launched. For each product we have recorded whether it was a success or failure, price charged for the product, marketing budget, competition price, and ten other variables.

(Answer(s))

- This is a regression problem. We are interested in inference. N = 500, d = 3 (profit, number of employees, industry)
- This is a classification problem. We are interested in prediction. N = 20, d = 13 (price charged, marketing budget, competition price, ten other variables)

Question 2:

(Question)

Think of real-life applications for machine learning

(a) Describe three real-life applications in which *classification* might be useful. Describe the target, as well as the features. Explain the application (inference/prediction).

(b) Describe three real-life applications in which *regression* might be useful. Describe the target, as well as the features. Explain the application (inference/prediction).

(Answer(s))

A.

(i) Spam detection
target: spam or not spam
features: subject, word counts, etc

(ii) Digit recognition
target: recognize digits from the image
features: number of strokes, aspect ratio, etc

(iii) Credit scoring
target: high risk or low risk
features: income, education, etc

B.

(i) Stock price prediction
target: stock price
features: past prices, annual growth, etc

(ii) Weather prediction
target: temperature
features: date, humidity, wind speed, etc

(iii) Housing price prediction
target: housing price
features: past prices, states, city, etc

Question 3:

(Question)

A university admissions office wants to predict the success of students based on their application material. They have access to past student records as training data.

(a) To formulate this as a supervised learning problem, identify a possible target variable. This should be some variable that measures success in a meaningful way and can be easily collected (in an automated manner) by the university. There is no one correct answer to this problem.

(b) Is the target variable continuous or discrete-valued?

(c) State at least one possible variable that can act as the predictor for the target variable you chose in part (a).

(d) Before looking at the data, would a linear model for the data be reasonable? If so, what sign do you expect the slope to be?

(Answer(s))

A.

A possible target variable could be the student's GPA at the time of graduation

B.

The target variable is continuous (from 0.00 to 4.00).

C.

One possible variable that can act as a predictor for the target variable is the SAT/ACT score.

D.

Yes, a linear model for the data would be possible. I expect the slope to be positive, with a positive correlation between SAT/ACT scores and GPA. Higher SAT/ACT scores might imply a higher GPA at the time of graduation.

Question 4:

(Question)

Suppose that we are given data samples (x_i, y_i):

x_i	0	1	2	3	4
y_i	0	2	3	8	17

(a) What are the population means, $\bar{x}$ and $\bar{y}$?

(b) What are the population variances and co-variances s_x^2, s_y^2 and s_{xy}?

(c) What are the least squares parameters for the regression line

$$y = w_0 + w_1 x + \epsilon.$$

(d) Using the linear model, what is the predicted value at $x = 2.5$?

(e) Compute the E_{in} = MSE = $(1/N)$ RSS.

(f) Calculate the R^2 (*confidence of determination*) and discuss the meaning of the number calculated.

(g) If one of the examples was changed to:

x_i	0	1	2	3	4
y_i	0	2	3	8	15

how does the value of the parameters change?

If the training data was changed further to:

x_i	0	1	2	3	4
y_i	0	2	3	8	9

how much would the parameters change (be qualitative(e.g. not very much increase, drastically decrease, etc...)?

(Answer(s))

A.

$$\overline{x} = \frac{0+1+2+3+4}{5} = 2$$

$$\overline{y} = \frac{0+2+3+8+17}{5} = 6$$

B.

$$s_{xx} = \frac{\sum_{i=1}^{5}(x_i - \overline{x})^2}{5} = 2$$

$$s_{yx} = \frac{\sum_{i=1}^{5}(y_i - \overline{y})(x_i - \overline{x})}{5} = 8$$

$$s_{yy} = \frac{\sum_{i=1}^{5}(y_i - \overline{y})^2}{5} = 37.2$$

C.

$$w_1 = \frac{s_{yx}}{s_{xx}} = \frac{8}{2} = 4$$

$$w_0 = \overline{y} - (w_1 * \overline{x}) = 6 - (4 * 2) = -2$$

D.

$$y_{\text{predicted}} = w_0 + (w_1 * x) = -2 + (4 * 2.5) = 8$$

E.

$$RSS = \sum_{i=1}^{5}(y_i - (w_0 + (w_1 * x_i)))^2 = 26$$

$$E_{\text{in}} = MSE = \frac{1}{N} * RSS = \frac{1}{5} * 26 = 5.2$$

F.

$$TSS = \sum_{i=1}^{5}(y_i - \overline{y})^2 = 186$$

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{26}{186} = 0.860215$$

R^2 represents the amount of variance in y that can be explained by the linear model used.

G.

If the original table was changed into the first example table, w_0 would increase slightly (-2 → -1.6) and w_1 would decrease slightly (4 → 3.6).
If the original table was changed into the second example table, w_0 would increase moderately (-2 → -0.4) and w_1 would decrease moderately (4 → 2.4).

Question 5:

(Question)

Using the examples in question 4 and starting with $w_0 = 0$ and $w_1 = 0$ perform 2 steps of gradient descent where the learning rate is $\alpha = 0.1$.

(Answer(s))

Step 1:

$$temp0 = w_0 - \frac{\alpha}{N} * \sum_{i=1}^{5} w_0 + (w_1 * x_i) - y_i$$

$$temp0 = 0 - \frac{0.1}{5} * (-30)$$

$$temp0 = 0.6$$

$$temp1 = w_1 - \frac{\alpha}{N} * \sum_{i=1}^{5} (w_0 + (w_1 * x_i) - y_i) * x_i$$

$$temp1 = 0 - \frac{0.1}{5} * (-100)$$

$$temp1 = 2$$

$$w_0 = 0.6,\ w_1 = 2$$

Step 2:

$$temp0 = w_0 - \frac{\alpha}{N} * \sum_{i=1}^{5} w_0 + (w_1 * x_i) - y_i$$

$$temp0 = 0.6 - \frac{0.1}{5} * (-7)$$

$$temp0 = 0.74$$

$$temp1 = w_1 - \frac{\alpha}{N} * \sum_{i=1}^{5} (w_0 + (w_1 * x_i) - y_i) * x_i$$

$$temp1 = 2 - \frac{0.1}{5} * (-33.4)$$

$$temp1 = 2.668$$

$$w_0 = 0.74,\ w_1 = 2.668$$

Question 6:

(Question)

A medical researcher wants to model, $z(t)$, the concentration of some chemical in the blood over time. She believes the concentration should decay exponentially in that

$$z(t) \approx z_0 e^{-\alpha t}, \tag{1}$$

for some parameters z_0 and α. To confirm this model, and to estimate the parameters z_0, α, she collects a large number of time-stamped samples $(t_i, z(t_i))$, $i = 1, \ldots, N$. Unfortunately, the model (1) is non linear, so she can't directly apply the linear regression formula.

(a) Taking logarithms, show that we can rewrite the model in a form where the parameters z_0 and α appear linearly.

(b) Using the transform in part (a), write the least-squares solution for the best estimates of the parameters z_0 and α from the data.

(c) Write a few lines of python code that you would compute these estimates from vectors of samples `t` and `z`.

(Answer(s))

A.

$$z(t) = z_0 e^{-\alpha t}$$
$$ln(z(t)) = ln(z_0 e^{-\alpha t})$$
$$ln(z(t)) = ln(z_0) + ln(e^{-\alpha t})$$
$$ln(z(t)) = ln(z_0) - \alpha t$$

B.

When comparing the above equation to the regression line ($y = w_0 + w_1 * x$), the following equivalencies are established:

$$y = ln(z)$$
$$w_0 = ln(z_0)$$
$$w_1 = -\alpha$$

$$x = t$$

Using these equivalencies on the best-estimates for w_0 and w_1, we get:

$$w_1 = \frac{s_{yx}}{s_{xx}} = \frac{\sum_{i=1}^{N}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

$$-\alpha = \frac{\sum_{i=1}^{N}(ln(z_i) - \overline{ln(z)})(t_i - \overline{t})}{\sum_{i=1}^{N}(t_i - \overline{t})^2}$$

$$\alpha = -\frac{\sum_{i=1}^{N}(ln(z_i) - \overline{ln(z)})(t_i - \overline{t})}{\sum_{i=1}^{N}(t_i - \overline{t})^2}$$

$$w_0 = \overline{y} - w_1 * \overline{x}$$

$$ln(z_0) = \overline{ln(z)} - w_1 * \overline{t}$$

$$z_0 = exp(\overline{ln(z)} - w_1 * \overline{t})$$

C.

```
import numpy as np
import math
y = np.log(z)
x = t
xm = np.mean(x)
ym = np.mean(y)
sxx = np.mean((x-xm)**2)
syy = np.mean((y-ym)**2)
sxy = np.mean((x-xm)*(y-ym))
w1 = sxy/sxx
w0 = ym - w1*xm
z0 = math.exp(w0)
alpha = -w1
```

Question 7:

(Question)

Consider a linear model of the form,

$$y \approx wx,$$

which is a linear model, but with the intercept forced to zero. This occurs in applications where we want to force the predicted value $\hat{y} = 0$ when $x = 0$. For example, if we are modeling $y =$ output power of a motor vs. $x =$ the input power, we would expect $x = 0 \Rightarrow y = 0$.

(a) Given data (x_i, y_i), write a cost function representing the residual sum of squares (RSS) between y_i and the predicted value $\hat{y}_i$ as a function of w.

(b) Taking the derivative with respect to w, find the w that minimizes the RSS.

(Answer(s))

A.

$$RSS = \sum_{i=1}^{N} (y_i - w * x_i)^2$$

B.

$$\frac{dRSS}{dw} = \sum_{i=1}^{N} (2)(-x_i)(y_i - w * x_i)$$

$$\sum_{i=1}^{N} (2)(-x_i)(y_i - w * x_i) = 0$$

$$\sum_{i=1}^{N} (x_i)(y_i - w * x_i) = 0$$

$$\sum_{i=1}^{N} x_i * y_i - \sum_{i=1}^{N} w * (x_i)^2 = 0$$

$$w * \sum_{i=1}^{N} (x_i)^2 = \sum_{i=1}^{N} x_i * y_i$$

$$w = \frac{\sum_{i=1}^{N} x_i * y_i}{\sum_{i=1}^{N} (x_i)^2}$$