

## HW2 — Regression

### 1. Linear Regression

1. Consider the data set  $x = \{1, 2, 3, 4, 5, 6, 10\}$ ,  $y = \{0, 1, 3, 2, 20, -6, 80\}$ .

- a) What are the sample means,  $\bar{x}$  and  $\bar{y}$ ?
- b) What are the sample variances and covariance,  $s_x^2$ ,  $s_y^2$ , and  $s_{xy}$ ?
- c) What is the solution (the values of  $a$  and  $b$ ) for the simple linear regression using the given data set with the function of the form

$$y = ax + b \quad (1-1)$$

(please show your calculation process step-by-step)?

- d) Is your predictor for (1-1) better than, worse than, or the same with the trivial predictor  $\hat{y} = \bar{y}$ ? Why?
- e) What is the solution (the value of  $k$ ) for the simple linear regression using the given data set with the function of the form

$$y = kx \quad (1-2)$$

(please show your calculation process step-by-step)?

- f) Is your predictor for (1-2) better than, worse than, or the same with the previous predictor of the form (1-1)? Why?
- g) What is the solution (the values of  $\alpha$ ,  $\beta$ , and  $\gamma$ ) for the regression using the given data set with the function of the form

$$y = \alpha x + \beta x^2 + \gamma \quad (1-3)$$

(please show your calculation process step-by-step)?

- h) Is your predictor for (1-3) better than, worse than, or the same with the previous predictor of the form (1-1)? Why?
- i) The programming solution and its performance using **sklearn** for (1-1) can be obtained as:

```
import numpy as np
from sklearn import linear_model
X = np.array([1, 2, 3, 4, 5, 6, 10]).reshape(-1,1)
Y = np.array([0, 1, 3, 2, 20, -6, 80])
""" fit y=ax+b """
model = linear_model.LinearRegression(fit_intercept=True)
model.fit(X, Y)
print ("a = %s, b=%s, score=%s" %
      (" {:.3 f}".format(model.coef_[0]),
       " {:.3 f}".format(model.intercept_),
       " {:.3 f}".format(model.score(X, Y))))
```

You can refer to the **sklearn** documentation ([link](#)) for more details and examples. Write a program to verify your solutions for (1-2) and (1-3). In your answer you should specify the key program (one or two lines of code) to obtain your solution, and whether the solution aligns with you previous answer for the respective functions ((1-2) and (1-3)).

j) Is there a unique solution for the regression problem using the function of the form

$$y = \beta_0 + \sum_{i=1}^{10} \beta_i x^i \quad (1-4)$$

with the given data set? If yes, what is your solution? If not, why isn't the solution unique? Use the `sklearn` package and the `linear_model` method to verify your answer. Is your programmed solution aligning with you answer above? If not, why?

## 2. Logistic Regression

1. Should the following be tackled as regression or classification problems? Think carefully and explain your answers.
  - a) Given a person's credit rating and zip code, predict how many children they will have.
  - b) Given a wildlife camera image, detect the number of legs on a creature.
  - c) Given a person's biometrics, predict whether they will live past the age of 80.
  - d) Given a person's biometrics, predict the person's age.

## 3. A "Bonus" Question, Again (1 pt)

From a scale of 1 to 5, how difficult is HW2? 1 is "I can do it in my sleep". 5 is "Bowen is ridiculous". 0 is "I refuse to answer this question". This is for my own reference to improve the quality of future assignments. Thanks!

*Submitted by Bowen Weng on .*