Linear Regression Worksheet

Work through this individually.

Question 1: Explain what $x_1, x_2, ..., y$ represent.

Question 2: Explain what $x_1^{(i)}, x_2^{(i)}, ..., y^{(i)}$ represent. What is the difference here?

Question 3: Using mathematical notation, explain the goal of supervised learning.

Question 4: What do we call the different functional forms in supervised learning and what are some different examples?

Question 5: Explain what parameters are? How does the choice of parameters affect the model performance?

Question 6: Explain the following assumption we might make about the relationship between our feature and target. You might want to draw this to help.

$$y = \alpha + \beta x_1$$

Question 7: We might model this assumed relationship with the following model. Why is it different to the above? **Hint:** This is quite subtle and skip if you are confused.

$$y^{(i)} = \alpha + \beta x_1^{(i)} + \varepsilon^{(i)}$$

Question 8: If $\alpha = 0.4$ and $\beta = 2$, draw a rough graph which includes 20 points and the linear regression we would hope to fit through this.

Question 9: [You may use a calculator or the internet to help solve this]. If $\alpha = 0.4$ and $\beta = 2$, answer the following questions.

- What is my predicted y ($\hat{y}$) when $x_1 = 14$?
- Describe the relationship between x_1 and y?
- When $x_1 = 0$, what is the predicted y?

Question 9: A student uses a linear regression model to estimate the effects of education (edu_i) and work experience (exp_i) on future wages (wage_i). They train their model on 100,000 people and get the following function. Note that wages are measures in thousands of US dollars ($), education is measured in years and work experience also in years.

$$\hat{\text{wage}}^{(i)} = 10 + 1.2\text{edu}^{(i)} + 0.8\text{exp}^{(i)}$$

Assuming that the relationship is casual, answer the following questions.

1. What wage would our model predict for someone with 10 years eduction and 20 years work experience? (2 marks)
2. How should we interpret the 1.2 in front of the edu_i variable? (2 marks)

3. How should we interpret the 10? (1 mark)

Question 10 [Extension]: Consider the following fitted linear regression which details the relationship between education and wages.

$$\hat{\text{wage}}^{(i)} = 10 + 1.2\text{edu}^{(i)} - 0.2\text{edu}^{(i)^2}$$

1. What do you notice about this regression which is different to the regression before?
2. What does this imply about the assumed relationship between education and wage
3. If the education is 1, what is the predicted wage?
4. If the education is 10, what is the predicted wage?
5. Plot a graph showing the implied relationship between education and wages.

Question 11 [Extension]:

1. What is differentiation?
2. Using differentiation, what is the rate of change in wage for a change in education in question 9?
3. What is the rate of change of wage for a change in education in question 10?
4. Explain the difference in your answer by thinking about how education might effect wages.

Question 12 [Extension]:

1. Think of another feature which might effect wage. What would the assumed linear relationship be and what might an example of a fitted regression look like?
2. [**Very hard**]: Differentiate wage with respect to both education and your new feature. What do you notice between how education effects wages and your new feature effect wages? Given this, you might you interpret β_1 and β_2?