# HMC CS 158, Fall 2017
# Problem Set 3 Exercises: Linear Regression

*Goals*:
- To investigate error for regression models.
- To extend a simple linear regression model using weights.

## Submission

You should submit any answers to the exercises in a single file `writeup.pdf`. This writeup should include your name and the assignment number at the top of the first page, and it should clearly label all problems. Additionally, cite any collaborators and sources of help you received (excluding course staff), and if you are using slip days, please also indicate this at the top of your document.

## 1 Linear Regression [4 pts]

Each plot in Figure 1 claims to represent prediction errors as a function of $x$ for a regression model trained on some dataset. (Note that we make no assumptions about the dataset.) Some of these plots could potentially be prediction errors for linear or quadratic regression models, but others could not. The regression models are trained with the least squares estimation criterion. Please indicate compatible models and plots. Justify your answer.

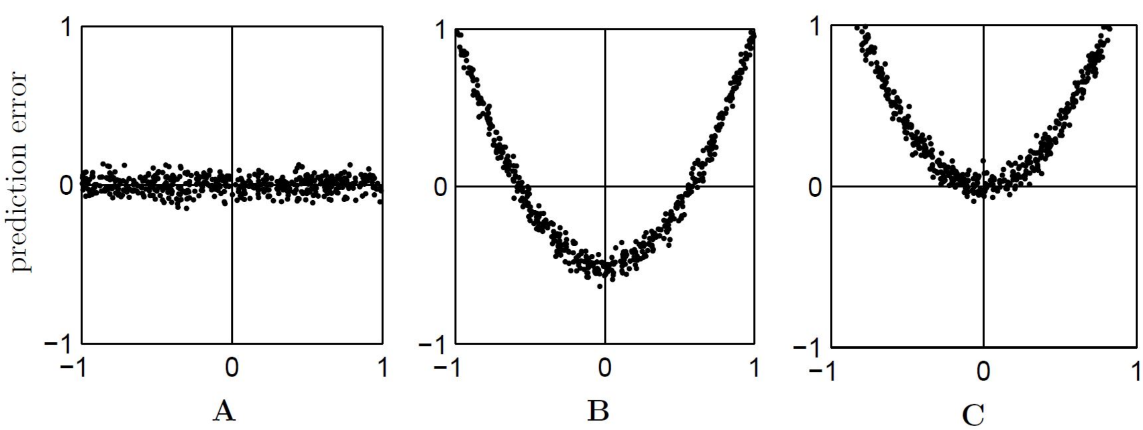


Figure 1: Possible prediction errors for trained regression models.

| | **A** | **B** | **C** |
|---|---|---|---|
| linear regression | ( ) | ( ) | ( ) |
| quadratic regression | ( ) | ( ) | ( ) |

Parts of this assignment are adapted from course material by Tommi Jaakola (MIT) and Andrew Ng (Stanford).

## 2 Locally Weighted Linear Regression [8 pts]

Consider a linear regression problem in which we want to "weight" different training examples differently. Specifically, suppose we want to minimize

$$J(\boldsymbol{\theta}) = \frac{1}{2}\sum_{i=1}^{n} w^{(i)} \left(\boldsymbol{\theta}^T \boldsymbol{x}^{(i)} - y^{(i)}\right)^2.$$

In class, we worked out what happens for the case where all the weights (the $w^{(i)}$'s) are the same. In this problem, we will generalize some of those ideas to the weighted setting.

(a) **(2 pts)** Let $\boldsymbol{X}$ and $\boldsymbol{y}$ be as defined in class. Show that $J(\boldsymbol{\theta})$ can also be written as

$$J(\boldsymbol{\theta}) = (\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})^T \boldsymbol{W} (\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})$$

for an appropriate diagonal matrix $\boldsymbol{W}$. State clearly what $\boldsymbol{W}$ is.

(b) **(3 pts)** If all the $w^{(i)}$'s equal 1, then we saw in class that the normal equation is

$$\boldsymbol{X}^T \boldsymbol{X}\boldsymbol{\theta} = \boldsymbol{X}^T \boldsymbol{y},$$

and that the value of $\boldsymbol{\theta}$ that minimizes $J(\boldsymbol{\theta})$ is given by $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$. By finding the derivative $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$ and setting that to zero, generalize the normal equation to this weighted setting, and give the new value of $\boldsymbol{\theta}$ that minimizes $J(\boldsymbol{\theta})$ in closed form as a function of $\boldsymbol{X}$, $\boldsymbol{W}$, and $\boldsymbol{y}$.

(c) **(3 pts)** Suppose we have a training set $\left\{(\boldsymbol{x}^{(i)}, y^{(i)})\right\}_{i=1}^{n}$ of $n$ independent examples, but in which the $y^{(i)}$'s were observed with differing variances. Specifically, suppose that

$$p\left(y^{(i)} | \boldsymbol{x}^{(i)}; \boldsymbol{\theta}\right) = \frac{1}{\sqrt{2\pi}\sigma^{(i)}} \exp\left(-\frac{\left(y^{(i)} - \boldsymbol{\theta}^T \boldsymbol{x}^{(i)}\right)^2}{2\left(\sigma^{(i)}\right)^2}\right)$$

i.e., $y^{(i)}$ has mean $\boldsymbol{\theta}^T\boldsymbol{x}^{(i)}$ and variance $(\sigma^{(i)})^2$ (where the $\sigma^{(i)}$'s are fixed, known constants). Show that finding the maximum likelihood estimate of $\boldsymbol{\theta}$ reduces to solving a weighted linear regression problem. State clearly what the $w^{(i)}$'s are in terms of the $\sigma^{(i)}$'s.