Homework 20
Dan Scarafoni

1.

a. Any linear function y = f(x) can be written in terms of two variable, m and be (see below)

$$y=mx+b$$

Thus, the equation can be written in vector form

$$\vec{w}=\langle w_{0,}\, w_1\rangle$$
$$\vec{x}=\langle w_x, x_1\rangle$$
$$y=\vec{x}\cdot\vec{y}$$

The values of W that suit any given equation are the ones that minimize the difference between the measured values in h(x) with their estimated counterparts in f(x), which is summarized in the equation below

$$w^{star}=argmin_w\left(\sum_x^1 (y_j-(w_1*x_j+w_0))^2\right)$$

b. A search is not needed as a closed form solution does exist for all satiable two dimensional solutions.

2.

a. we can assume that the function for batting average can be expressed in terms of all variables that influence the batting average X.

$$y=B(X)$$
$$given\{(x_{1,}\, y_1),(x_{2,} y_2)\ldots(x_n, y_n)\}$$

The hypothesis space H for the answer to the question is expression below, assuming that X is the set of all variables that have direct influence on the batting average

$$H(x)=\sum_i w_i*x_j, i$$
$$where:$$
$$w_i=coefficient\ \ for\ \ term\ \ i$$
$$x_j=a\, subterm\ \ function\ \ of\ \ x$$

thus, an answer can be calculated in a similar manner to question one

$$w^{star}=argmin_w\left(\sum_j L_2(y_j, w\cdot x_j)\right)$$

No closed form solution exists for multivariate regressions, and as such a search is needed. Gradient search, a form of local search, can be used to approximate the best fit solution. The equation for each weight is below.

$$w_i \leftarrow w_i+a\sum_j x_{j,i}(y_j-h_w(\boldsymbol{x_j}))$$

b. It is sometimes wise. Because multivariate equations can utilize an arbitrary number of variables, overfitting is a real possibility. Regularization provides a means of avoiding this. Regularization can be calculated in terms of empirical loss and complexity of a hypothesis equation, as show below

$$Cost(h_w) = EmpLoss(h) + \lambda \sum_i |w_i|^q$$
$$where:$$
$$q = degree\ of\ polynomial\ equation\ sums\ minimzed\ by\ the\ regularization\ function$$

c. By using the gradient descent algorithm specified above, an answer can be calculated through continued iteration through the multidimensional hypothesis space. The following equation (repeated below) describes the process.

$$w_i \leftarrow w_i + a \sum_j x_{j,i}(y_j - h_w(\mathbf{x_j}))$$

3.

a. a decision boundary is a line or surface that separates two classes of information in a graph comparing two or more classes of data.

b. A linear separator is decision boundary that is a linear equation

c. linearly separable is a term that describes two classes of data that can be separated by a linear separator

4. Linear regression attempt to find an equation that best describes and a set of data points. It attempts to describe the general relationship between two variables. Classification may use a linear regression in order to divide data into select groups (as a decision boundary). Depending on their values with respect to the regression.

5. Logistic regression creates a continuous threshold in which the hypothesis is non-differentiable. In previous examples, the data could only be on one side of the threshold, and was classified with either the value 0 or 1. By using a continuous, differentiable function instead, these problems can be mitigated. The binary values of 0 and 1 can be replaced with a spectrum of probability, a “soft barrier” in contrast to the linear regression's “hard barrier.”

There are downsides to this method, however, there are no easily calculable closed-form solutions for the value of **w** when using logistic regression.