MAT 345 - Homework 6

Due Wednesday, November 7, 2018, in class

1. (1 point) Suppose we run linear regression on a data set $\mathcal{D}$ and find a weight vector $\mathbf{w}$ for which $E_{\text{in}}(\mathbf{w}) = 0$. What does this say about the data set $\mathcal{D}$?

2. (5 points). Consider the following data set

x	6	-2	4.8	2	-0.8	-1	2	6	5.8	5	-1.5	0
y	0	0	2.8	4	2.8	-2.8	-4	0.6	1.4	-2.8	-2	3.5

(a) Draw a scatter plot for this data to convince yourself that the points are approximated by a circle with center $(a, 0)$ for some a.

(b) Suppose the circle from (a) has center $(a, 0)$ and radius R. Then each data point (x, y) satisfies

$$(x-a)^2 + y^2 \approx R^2 \quad \Rightarrow \quad y^2 \approx -x^2 + 2a\,x - a^2 + R^2.$$

Note that we cannot use linear regression at this time. However, using a transformation, as discussed in class, will allow us to use regression as follows.

-let $y_2 = y^2$ for each data point.
-let $x_0 = 1$ for each data point.
-let $x_1 = x$ for each data point.
-let $x_2 = x^2$ for each data point.

Re-write the given data in this notation:

x_0												
x_1	6	-2	4.8	2	-0.8	-1	2	6	5.8	5	-1.5	0
x_2												
y_2												

(c) Use linear regression to find the weight vector $\mathbf{w} = [w_0, w_1, w_2]^T$ that best approximates

$$y_2 \approx w_0 x_0 + w_1 x_1 + w_2 x_2 = \mathbf{w}^T\mathbf{x}$$

(d) Write the equation of the circle observed in (a).

3. (4 points) Use logistic regression to find the weight vector $\mathbf{w}$ that best approximates the probability

$$P(y\,|\,\mathbf{x}) \approx \theta(y\,\mathbf{w}^T\mathbf{x}) = \frac{e^{y\mathbf{w}^T\mathbf{x}}}{1 + e^{y\mathbf{w}^T\mathbf{x}}}$$

for the following data set:

x_1	32	45	60	53	25	68	82	38	67	92	72	21	26	40	33	45	61	16	18	22
x_2	3	2	2	1	4	1	2	5	2	2	3	5	3	4	3	1	2	3	4	6
y	-1	-1	1	-1	-1	1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	1	-1