Julio Soldevilla
EECS 545 Winter 2018 — Problem Set 2

Problem 1 Problem 1

Proof:

1. For the first part of this problem, for the linear regression we get a test error of 0.4996621 and we get a the following plot predicting the labels for the test instances:

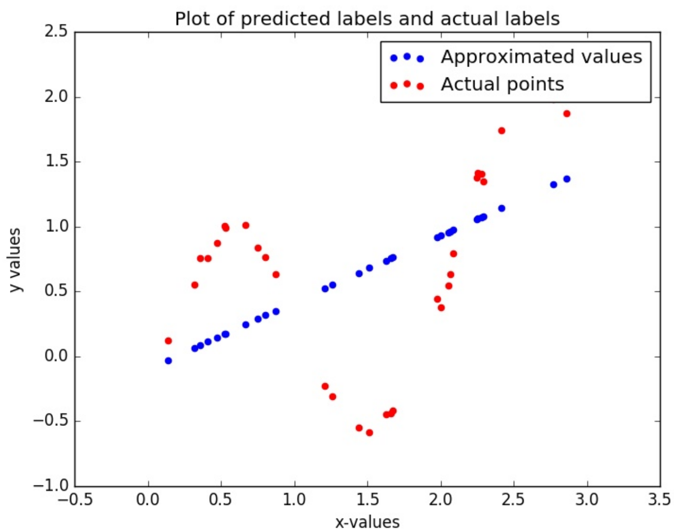

Figure 1: Image showing the plot of the prediction using linear regression

2. For the second part of this problem, we use locally weighted linear regression. In this part of the problem, we found that when $\tau = 0.2$, we get a test error of 0.01395202 and when $\tau = 2$, we get a test error of 0.4253173. Also, we present plots showing the prediction lables with the actual labels for the test instances when $\tau = 0.2$ (first figure below) and when $\tau = 2$ (second figure below).

■

Problem 2 Problem 2

Proof:

1. This part of the problem is in the code submitted to canvas.

2. In this part of the problem, we found that the marginal distribution $P(X_1)$ is a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. The plot of this distribution is the following one:

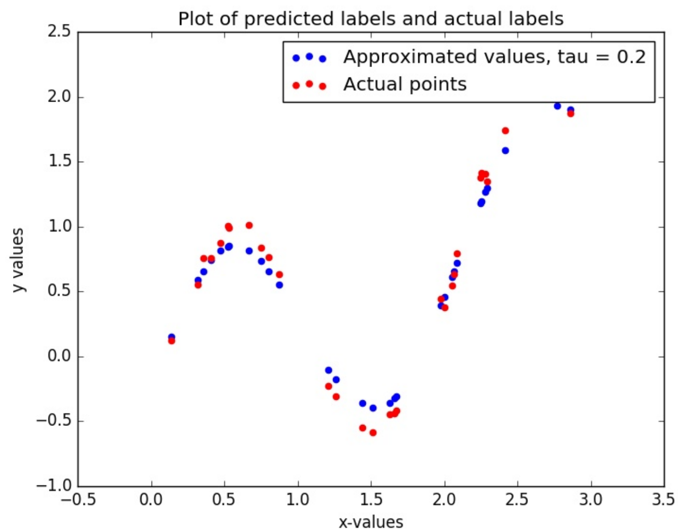

Figure 2: **Problem 1b:** Image showing the plot of the prediction using locally weighted linear regression and $\tau = 0.2$

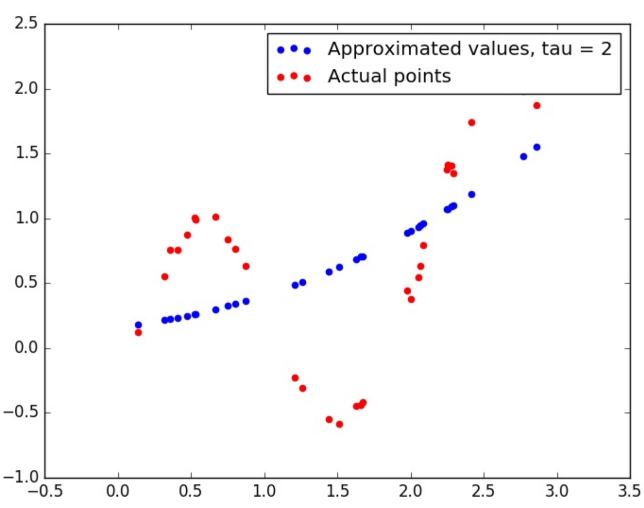

Figure 3: **Problem 1b:** Image showing the plot of the prediction using locally weighted linear regression and $\tau = 2$

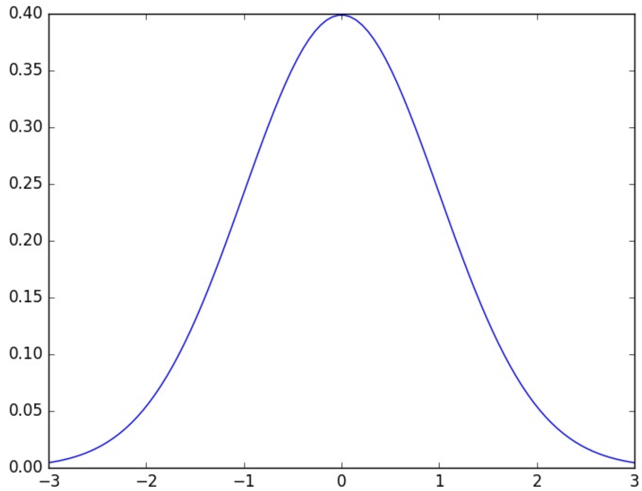

Figure 4: **Problem 2a:** Image showing hte marginal distribution $P(X_1)$

3. The solution of this problem is in the code submitted to canvas.

4. In this part of the problem, we found that the marginal distribution $P(X_1, X_4|X_2 = 0.1, X_3 = -0.2)$ follows a normal distribution with mean $\mu = \begin{bmatrix} 0.55 & 0.15 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} 0.75 & -0.75 \\ -0.75 & 1.75 \end{bmatrix}$. The heat map of this distribution is shown below:

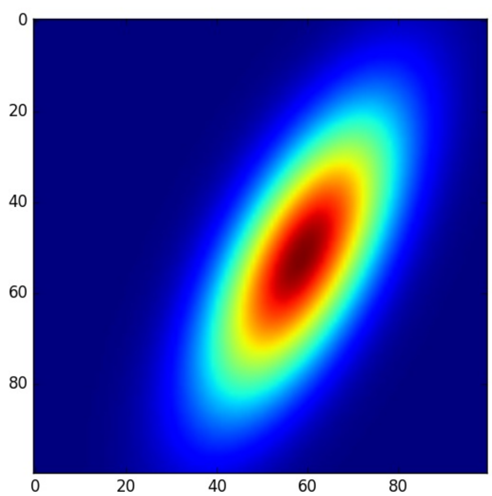

Figure 5: **Problem 2b:** Image showing the conditional distribution $P(X_1, X_4|X_2 = 0.1, X_3 = -0.2)$

■

Problem 3 Problem 3

Proof: Our solution for this problem is presented in the following list, which presents the mean and covariance matrices for a particular instance:

1. This is the mean for initial distribution of w $\begin{bmatrix} 0.31207185 & -0.38444814 \end{bmatrix}$ This is the covariance matrix for initial distribution of w $\begin{bmatrix} 0.40995687 & 0.1109261 \\ 0.1109261 & 0.36334771 \end{bmatrix}$

2. This is the mean for the distribution of w after 1 instance(s) $\begin{bmatrix} 0.31207185 & -0.38444814 \end{bmatrix}$ This is the covariance matrix for the distribution of w after 1 instance(s) $\begin{bmatrix} 0.40995687 & 0.1109261 \\ 0.1109261 & 0.36334771 \end{bmatrix}$

3. This is the mean for the distribution of w after 10 instance(s) $\begin{bmatrix} 0.97746676 & -0.52199623 \end{bmatrix}$ This is the covariance matrix for the distribution of w after 10 instance(s) $\begin{bmatrix} 0.06639281 & -0.02378635 \\ -0.02378635 & 0.0918552 \end{bmatrix}$

4. This is the mean for the distribution of w after 20 instance(s) $\begin{bmatrix} 0.43026927 & -0.22064332 \end{bmatrix}$ This is the covariance matrix for the distribution of w after 20 instance(s) $\begin{bmatrix} 0.01792457 & -0.00630496 \\ -0.00630496 & 0.04767232 \end{bmatrix}$

The plots for each of the four cases above are presented below:

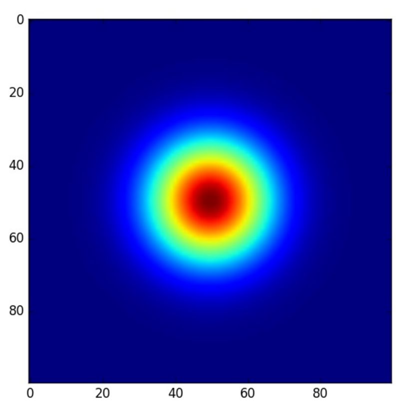

Figure 6: **Problem 3:** Image showing the initial distribution of w

■

Problem 4 Problem 4

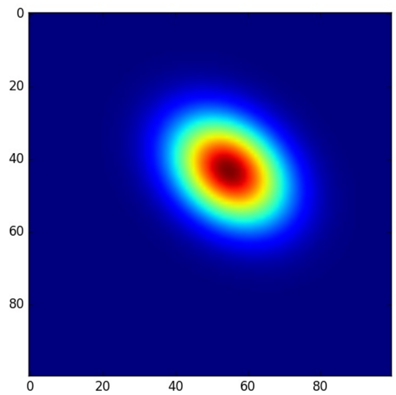

Figure 7: **Problem 3:** Image showing the distribution of w after 1 instance

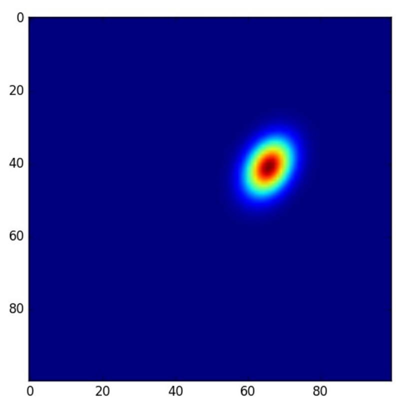

Figure 8: **Problem 3:** Image showing the distribution of w after 10 instance

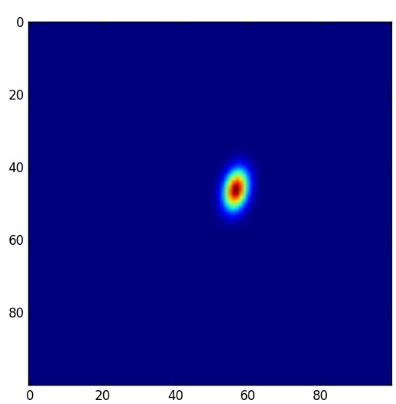

Figure 9: **Problem 3:** Image showing the distribution of w after 20 instances

Proof:

1. Part a is divide into two smaller parts

   (a) The joint distribution over $y(x_1), ..., y(x_n)$ is Gaussian distribution, with mean $\mu = \vec{0}$ where μ has size $n \times 1$ and covariance matrix
   $$\Sigma_n = \begin{bmatrix} exp^{-\frac{\|x_1-x_1\|^2}{2\sigma^2}} & exp^{-\frac{\|x_1-x_2\|^2}{2\sigma^2}} & \dots & exp^{-\frac{\|x_1-x_n\|^2}{2\sigma^2}} \\ exp^{-\frac{\|x_2-x_1\|^2}{2\sigma^2}} & exp^{-\frac{\|x_2-x_2\|^2}{2\sigma^2}} & \dots & exp^{-\frac{\|x_2-x_n\|^2}{2\sigma^2}} \\ \vdots & \dots & \dots & \vdots \\ exp^{-\frac{\|x_n-x_1\|^2}{2\sigma^2}} & exp^{-\frac{\|x_n-x_2\|^2}{2\sigma^2}} & \dots & exp^{-\frac{\|x_n-x_n\|^n}{2\sigma^2}} \end{bmatrix}$$

   (b) The following figures are the plots for the sampling form the GP prior shown above. Each figure corresponds to one value of $\sigma = \{0.3, 0.5, 1\}$

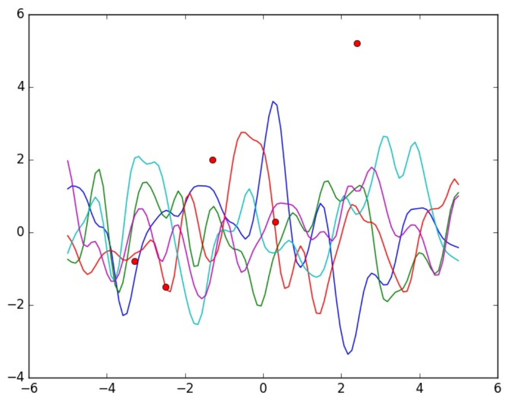

Figure 10: **Problem 4:** Image showing the 5 samples from the GP with $\sigma = 0.3$

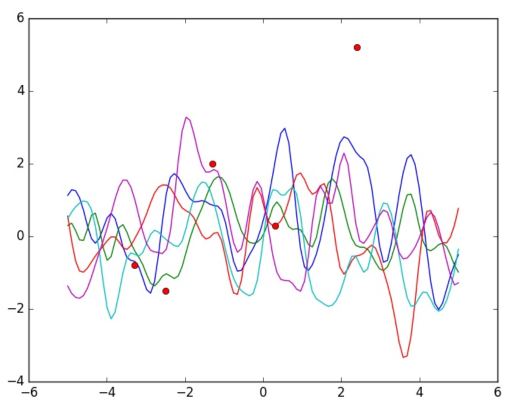

Figure 11: **Problem 4:** Image showing the 5 samples from the GP with $\sigma = 0.5$

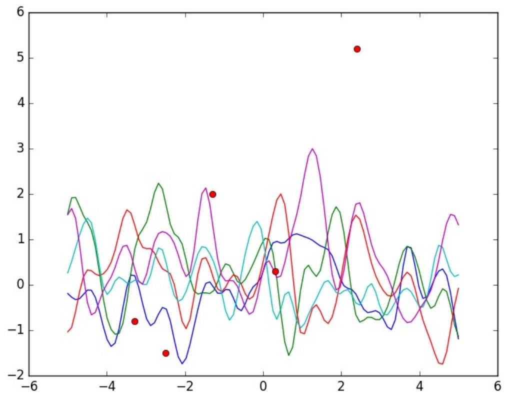

Figure 12: **Problem 4:** Image showing the 5 samples from the GP with $\sigma = 1$

2. Part b is also divided in 2 parts:

   (a) This time, again the posterior follows a Gaussian distribution, with similar mean and sigma as above.

   (b) The following figures are the plots for the sampling form the GP posterior shown above. Each figure corresponds to one value of $\sigma = \{0.3, 0.5, 1\}$

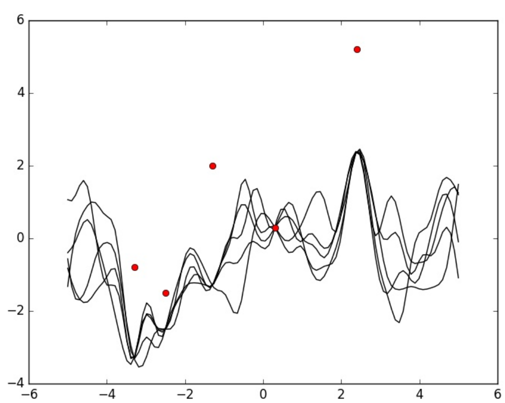

Figure 13: **Problem 4:** Image showing the 5 samples from the posterior GP with $\sigma = 0.3$

■

Problem 5 Problem 5

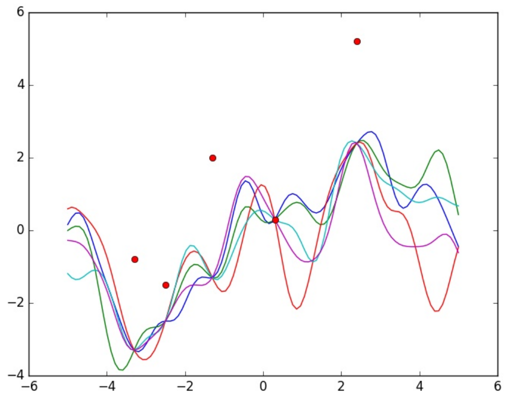

Figure 14: **Problem 4:** Image showing the 5 samples from the posterior GP with $\sigma = 0.5$

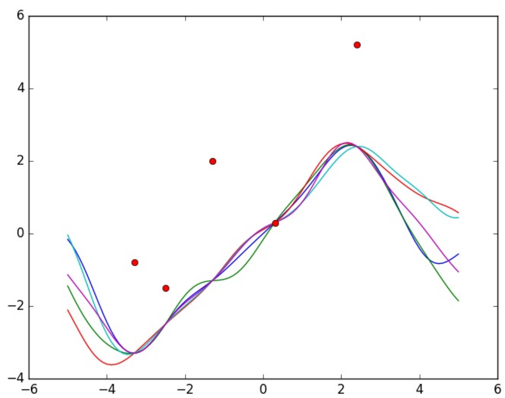

Figure 15: **Problem 4:** Image showing the 5 samples from the posterior GP with $\sigma = 1$

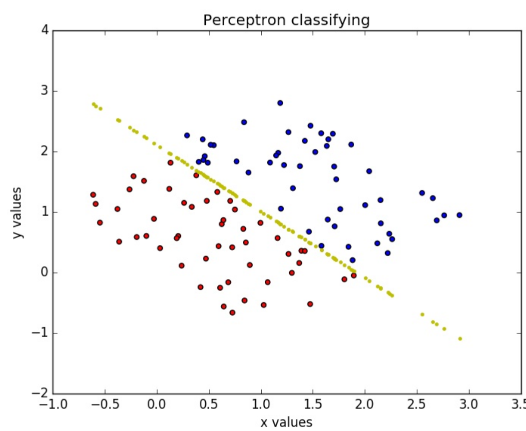

Figure 16: **Problem 5:** Image showing the linear hyperplane dividing the points found usng only the perceptron update rule

Proof: For this problem, for **part a)**, we present a plot that shows the linear hyperplane dividing the points we found using only the perceptron update rule. This is the shown in the followin picture.

For **part b)** of the problem, we do the same analysis for a set of data that doesn't have a linear separation. IN this case, we present the linear separation found by the perceptron update algorithm plotted with yellow points in the figure below. We can see that this line is not as good as it can be. To improve we apply also the pocket algorithm, which implies that we will store in memory the vector of w that produces the highest correct predictions throughout the algorithm. If the vector w that the algorithms ends with has a higher matching rate, then we use an output of the algorithm the last w. If the vector w we were saving throughout the algorithm is better, then we take as an output the saved vector w. In this way, we make sure we have the vector of coefficient that produces the most matches. In the figure below, the line in black is obtained using the vector w from the pocket algorithm and we can see that it does a better job of classifying the points.

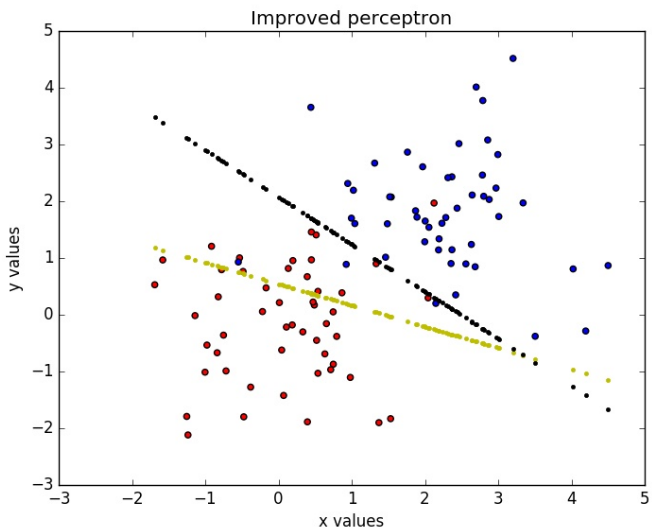

Figure 17: **Problem 3:** Image showing linear hyperplane dividing the points found using only the perceptron update rule and the pocket algorithm. The yellow line is obtained using only the perceptron update rule and the other black line is obtained using the pocket algorithm

■

Problem 6 Problem 6

Proof: For this problem we found that the weight vector without bias term is: $[0.49230156, -1.02181685, -0.38697494, -0.70922472, -0.02087894, 0.14882702, -0.07118593, -0.09869681, 0.15392742, 0.58795622, -0.05514312, 0.82930682, -0.62670417, 0.32124253, -0.79053791, 0.55524013, -0.44701531, -0.75471243, -0.26418911, -0.21516875, -0.9294681, 0.03643845, -0.11790563, -0.26960344, -0.42819214, 0.35976133, -0.90088309, -1.05377673, -0.54921289, 0.60597135]$

Also, for this problem, this is the final training cross entropy: 0.193449363355, this is the final test cross entropy: 0.181117073447, this is the final training classification accuracy percentage: 93.43832021% and this is the final test classification accuracy percentage: 94.1489361702%

Finally, the plots for the training and test classification accuracy vs number of iterations (first picture) and the plot of average training and test erros vs number of iterations (second picture) are shown below.

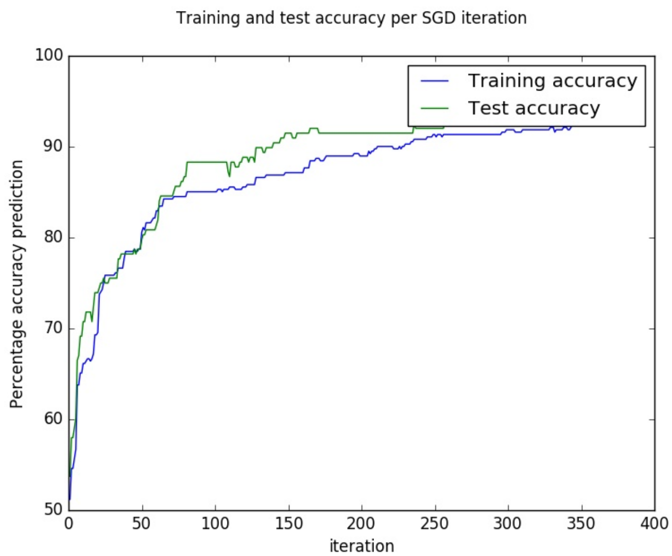

Figure 18: **Problem 6**: Image showing the training and test classification accuracy vs number of iterations plot

■

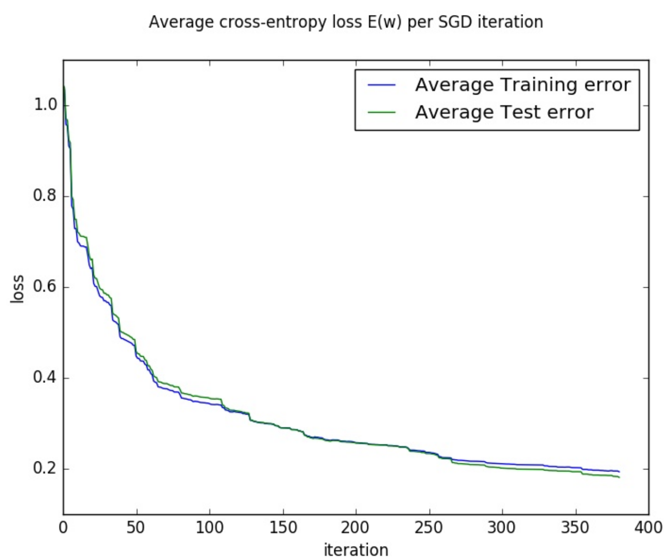

Figure 19: **Problem 6**: Image showing the training and test cross entropy vs number of iterations plot