Homework Assignment 2

COGS 181: Neural Networks and Deep Learning

Due: Oct. 15, 2017, 11:59pm
Instructions: Please answer the questions below, attach your code, and insert figures to create a pdf file; submit your file to TED (ted.ucsd.edu) by 11:59pm, 10/15/2017. You may search information online but you will need to write code/find solutions to answer the questions yourself.

Late Policy: %5 of the total points will be deducted on the first day past due. Every 10% of the total points will be deducted for every extra day past due.

System Setup: You are free to choose either pip or anaconda as the package installer. After the installation of one of the installer, type pip/conda install $PACKAGE_NAME in the terminal to install python packages.

Grade: ____ out of 100 points

1 (6 Points) Vector Norm

L1-norm is given by $||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|$ and L2-norm is given by $||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$. Compute the L1-norm and L2-norm of the following vector:

(1) $\begin{bmatrix} -1 \\ -5 \\ -3 \end{bmatrix}$

(2) $\begin{bmatrix} 0 \\ 4 \\ 16 \end{bmatrix}$

2 (6 Points) Vector Projection

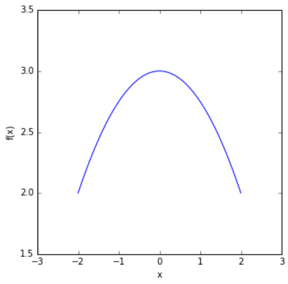
What is the cosine value between the following vectors:

(1) $\begin{bmatrix} 4 \\ -5 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 8 \\ 10 \\ 6 \end{bmatrix}$

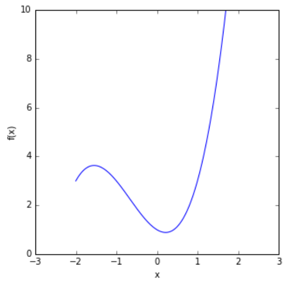
(2) $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$

3 (16 Points) Convex Functions

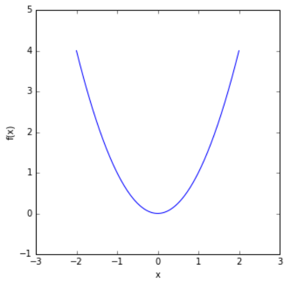
Please identify the convexity for the following functions. Simply write down whether the function is convex or non-convex.



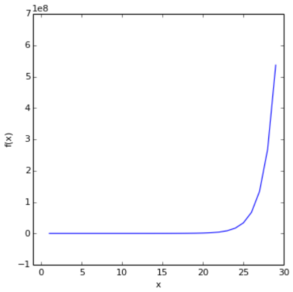
(a)



(b)



(c)



(d)

4 (24 points) Error Metrics

In the table below, we have the specifications of five computers and we have trained a classifier to differentiate whether a computer's price is high or low. Assuming that the classifier can be formulated as

$$y_i = \begin{cases} +1, & < W, x_i > +b \geq 0 \\ -1, & < W, x_i > +b < 0. \end{cases}$$

, where W is the weight of each feature, x_i is the i_{th} instance of data and b is the bias. Please answer the following question. Given: $W = \begin{bmatrix} 1.2 \\ 2 \\ 0.5 \\ 0.7 \end{bmatrix}$ $b = -20$

Specifications	CPU Cores	GPU RAM	RAM	Hard Disk	Price
Computer 1	8	8GB	16GB	4TB	$High$
Computer 2	4	1GB	1GB	16TB	Low
Computer 3	6	4GB	4GB	2TB	$High$
Computer 4	4	2GB	4GB	1TB	Low
Computer 5	8	4GB	8GB	2TB	$High$

1. Please write down the data matrix X and label matrix Y. For label matrix Y, please assign 1 to label $High$ and -1 to label Low.

2. Compute and write down $< W, x_i > +b$ for each x_i and their predicted label in $+1$ and -1 based on the classification rule.

3. Compute the performance of this classifier in terms of accuracy, recall, precision and f-score.

5 (24 points) Polynomial Regression

Now use the same data from Homework 1 Q6, but learn a polynomial regressor such as $y = ax^2 + bx + c$ to fit the data.

(1) Derive the formulation for the second order least square problem starting from L2-loss. Hint: You can start by writing the L2 loss in matrix form and set the derivative with respect to W to 0. The X in the code from homework 1 Q6 should contain an x^2 term now, for example

```
X = [ones(length(x),1) x^1 x^2].
```

(2) Write your code to compute the derived polynomial regressor $y = ax^2 + bx + c$ using the given data (data.txt), and overlay it on the data point in the same figure as in Homework 1 Q6. Please paste your code and figures in the report for this question.

(3) Explain which formulation (line or polynomial) is a better model here with respect to fitting this dataset. You can compute the L2 distance between your prediction $\hat{y}$ and target y to justify your choice.

6 (24 points) L1 Loss Regression

In this problem we will be using data_2.txt. You can load the file using the code that is similar to homework 1 Q6.

(1) The method we implemented in homework 1 is based on L2 loss. L1 loss is also frequently used in regression tasks. To implement L1 loss method, please first derive the gradient of the L1 loss w.r.t. W for data instances $x_i, i = 1..n$. The L1 loss is given by $\sum_{i=1}^{n} |y_i - x_i^T W|$.

(2) Implement the line (first order regressor) formulation with L1 loss, which is given by $\sum_{i=1}^{n} |y_i - x_i^T W|$, via gradient descent with 2000 iterations of updates and alpha = 0.00001. Please paste your code in the report for this question.

(3) Plot the figure with the original data points from (data_2.txt), the fitted line from (2) (using L1 loss) and the fitted line from homework 1 Q6(using L2 loss).