CS416-23 – HW1

David Montenegro

Introduction

Please complete all exercises and problems below.

- All the files can be found in

  `http://www.cs.wm.edu/~liqun/teaching/cs416/hw1/`

- You can also copy to your directory on a department machine by:

  `cp ~liqun/public_html/teaching/cs416/hw1/* .`

Your submission consists of four steps:

1. Create hw1.pdf with your solutions to the following problems. The solutions can be typed or written and scanned but the resulting pdf must be high quality and easily readable. Put your name in the file.
2. You'll need to create or edit these files in the directory hw1. Complete the requested code in these files.
   - exercise_1.ipynb
   - gd.py
3. Compile your exercise_1.ipynb into exercise_1.pdf
4. Submit the following files:

   `hw1.pdf  exercise_1.pdf  exercise_1.ipynb  gd.py`

1 SVM (5 points + 10 points)

1. Suppose you are given the following training data

positive: (1,2,3) (1,1,4)
negative: (3,2,-1) (4,3,-2) (3,5,-3)

Write down the SVM optimization for those training data including the optimization objective and the constraints.

Minimize $\frac{1}{2}w^T w = \frac{1}{2}(a_1 * a_1 + a_2 * a_2 + a_3 * a_3)$

$(-1) * [a_1 * 3 + a_2 * 2 + a_3 * (-1) + b] \geq 1$

$1 * [a_1 * 1 + a_2 * 2 + a_3 * 3 + b] \geq 1$

$(-1) * [a_1 * 4 + a_2 * 3 + a_3 * (-2) + b] \geq 1$

$1 * [a_1 * 1 + a_2 * 1 + a_3 * 4 + b] \geq 1$

$(-1) * [a_1 * 3 + a_2 * 5 + a_3 * (-3) + b] \geq 1$

5 constraints corresponding to 5 data points

2. Suppose you are given the following training data

positive: (1,2) (1,1) (2,1) (0,1)
negative: (3,2) (4,3) (3,5)

Which points are support vectors? What is the decision boundary if you use SVM? In this problem, you can simply look at the points and decide which points are support vectors and then calculate the decision boundary.

$(1,2)$, $(2,1)$, and $(3,2)$ are support vectors

$(1,2)$ $(2,1)$

Slope: $\frac{(1-2)}{(2-1)} = -1$

$y = (-1) * x + b$

$2 = (-1) * 1 + b \rightarrow b = 3$

$(3,2)$

$y = (-1) * x + 5$

$y = (-1) * x + 4 \rightarrow x_2 + x_1 - 4 = 0$

Decision boundary: $x_2 + x_1 - 4 = 0$

2 Decision Tree (20 points)

Consider the following dataset consisting of five training examples followed by three test examples:

x1	x2	x3	y
-	+	+	-
+	+	+	+
-	+	-	+
-	-	+	-
+	+	-	+
+	-	-	?
-	-	-	?
+	-	+	?

There are three attributes (or features or dimensions), x1, x2 and x3, taking the values + and -. The label (or class) is given in the last column denoted y; it also takes the two values + and -.

Simulate each of the following learning algorithms on this dataset. In each case, show the final hypothesis that is induced, and show how it was computed. Also, say what its prediction would be on the three test examples.

- The decision tree algorithm discussed in class. For this algorithm, use the information gain (entropy) impurity measure as a criterion for choosing an attribute to split on. Grow your tree until all nodes are pure, but do not attempt to prune the tree.

  3+, 2− : E = 0.971

  $x1$

  $+$: 2+, 0−, E = 0 | $-$: 1+, 2−, E = 0.918

  Gain = $0.971 - (\frac{2}{5} * 0 + \frac{3}{5} * 0.918) = 0.420$

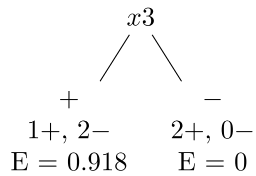
  3+, 2− : E = 0.971

  $x2$

  $+$: 3+, 1−, E = 0.811 | $-$: 0+, 1−, E = 0

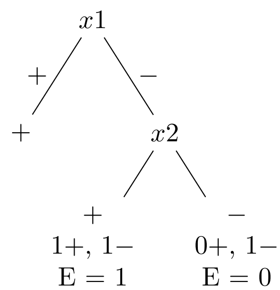
  Gain = $0.971 - (\frac{4}{5} * 0.811 + \frac{1}{5} * 0) = 0.322$

$3+, 2-$: E = 0.971



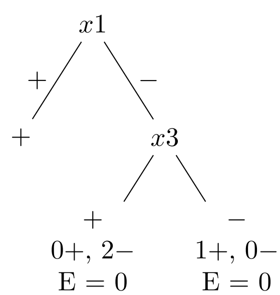
$$\text{Gain} = 0.971 - (\tfrac{3}{5} * 0.918 + \tfrac{2}{5} * 0) = 0.420$$

$1+, 2-$: E = 0.918

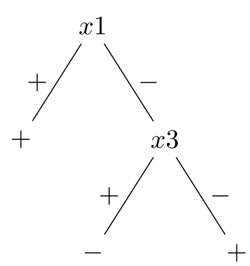


$$\text{Gain} = 0.918 - (\tfrac{2}{3} * 1 + \tfrac{1}{3} * 0) = 0.251$$

$1+, 2-$: E = 0.918



$$\text{Gain} = 0.918 - (\tfrac{2}{3} * 0 + \tfrac{1}{3} * 0) = 0.918$$



Prediction on test examples:

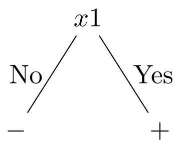
Example 1: +

Example 2: +

Example 3: +

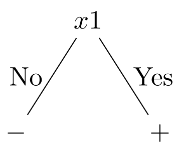
- AdaBoost. For this algorithm, you should interpret label values of + and - as the real numbers +1 and -1. Use decision stumps as weak hypotheses, and assume that the weak learner always computes the decision stump with minimum error on the training set weighted in AdaBoost algorithm. Note that a decision stump is a one-level decision tree. Run your boosting algorithm for three rounds and list the intermediate results.

  $x1 > 0$

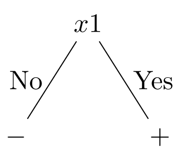
  Error $= \frac{1}{5}$

  $x1 < 2$

  Error $= \frac{2}{5}$

  $x1 > -2$

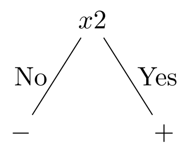
  

  Error $= \frac{2}{5}$

  h1 $= (x1 > 0)$
  error $= \frac{1}{5}$
  beta $= \dfrac{(1 - \frac{1}{5})}{(\frac{1}{5})} = 4$
  alpha = log(beta) = log(4) = 1.386
  $\frac{1}{5} * 4 + \frac{4}{5} * 1 = \frac{8}{5}$

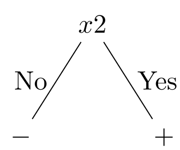
	1	2	3	4	5
weight	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
reweight	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{4}{5}$	$\frac{1}{5}$	$\frac{1}{5}$
rescale	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$

$x2 > 0$



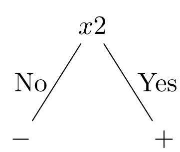
Error $= \frac{1}{8}$

$x2 < 2$



Error $= \frac{2}{8} = \frac{1}{4}$

$x2 > -2$



Error $= \frac{2}{8} = \frac{1}{4}$

h2 $= (x2 > 0)$

error $= \frac{1}{8}$

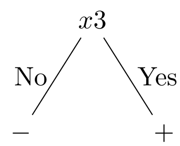
beta $= \dfrac{(1 - \frac{1}{8})}{(\frac{1}{8})} = 7$

alpha = log(beta) = log(7) = 1.946

$\frac{7}{8} * 1 + \frac{1}{8} * 3 + \frac{1}{2} * 1 = \frac{7}{4}$

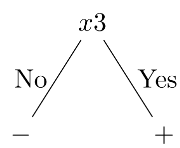
	1	2	3	4	5
weight	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$
reweight	$\frac{7}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$
rescale	$\frac{1}{2}$	$\frac{1}{14}$	$\frac{2}{7}$	$\frac{1}{14}$	$\frac{1}{14}$

$x3 < 0$



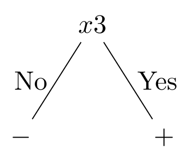
Error $= \frac{1}{14}$

$x3 < 2$



Error $= \frac{8}{14} = \frac{4}{7}$

$x3 > -2$



Error $= \frac{8}{14} = \frac{4}{7}$

h2 $= (x3 < 0)$
error $= \frac{1}{14}$
beta $= \dfrac{(1 - \frac{1}{14})}{(\frac{1}{14})} = 13$
alpha $=$ log(beta) $=$ log(13) $= 2.565$
$\frac{1}{2} * 1 + \frac{13}{14} * 1 + \frac{2}{7} * 1 \frac{1}{14} * 2 = \frac{13}{7}$

	1	2	3	4	5
weight	$\frac{1}{2}$	$\frac{1}{14}$	$\frac{2}{7}$	$\frac{1}{14}$	$\frac{1}{14}$
reweight	$\frac{1}{2}$	$\frac{13}{14}$	$\frac{2}{7}$	$\frac{1}{14}$	$\frac{1}{14}$
rescale	$\frac{7}{26}$	$\frac{1}{2}$	$\frac{2}{13}$	$\frac{1}{26}$	$\frac{1}{26}$

H(X) = 1.386*h1+1.946*h2+2.565*h3 = $1.386 * (x1 > 0) + 1.946 * (x2 > 0) + 2.565 * (x3 < 0)$

Prediction on test examples:

Example 1: $1.386 * (1) + 1.946 * (-1) + 2.565 * (1) = 2.005$, +

Example 2: $1.386 * (-1) + 1.946 * (-1) + 2.565 * (1) = -0.767$, $-$

Example 3: $1.386 * (1) + 1.946 * (-1) + 2.565 * (-1) = -3.125$, $-$

3 Naive Bayes (10 points)

The following table contains training examples that help predict whether a person is likely to have come kind of disease.

ID	PAIN?	MALE?	SMOKES?	WORK OUT?	DISEASE?
1.	yes	yes	no	yes	yes
2.	yes	yes	yes	no	yes
3.	no	no	yes	no	yes
4.	no	yes	no	yes	no
5.	yes	no	yes	yes	yes
6.	no	yes	yes	yes	no
7.	no	yes	yes	no	?

Use Naive Bayes method to predict whether the last person will have the disease. Be sure to use Laplace smoothing. Show the steps for your calculation.

P(disease) = $\frac{4}{6} = \frac{2}{3}$

P(no disease) = $\frac{2}{6} = \frac{1}{3}$

P(pain|disease) = $\frac{3+1}{4+1(2)} = \frac{4}{6} = \frac{2}{3}$

P(no pain|disease) = $\frac{1+1}{4+1(2)} = \frac{2}{6} = \frac{1}{3}$

P(pain|no disease) = $\frac{0+1}{2+1(2)} = \frac{1}{4}$

P(no pain|no disease) = $\frac{2+1}{2+1(2)} = \frac{3}{4}$

P(male|disease) = $\frac{2+1}{4+1(2)} = \frac{3}{6} = \frac{1}{2}$

P(no male|disease) = $\frac{2+1}{4+1(2)} = \frac{3}{6} = \frac{1}{2}$

P(male|no disease) = $\frac{2+1}{2+1(2)} = \frac{3}{4}$

P(no male| no disease) = $\frac{0+1}{2+1(2)} = \frac{1}{4}$

P(smokes|disease) = $\frac{3+1}{4+1(2)} = \frac{4}{6} = \frac{2}{3}$

P(no smokes|disease) = $\frac{1+1}{4+1(2)} = \frac{2}{6} = \frac{1}{3}$

P(smokes|no disease) = $\frac{1+1}{2+1(2)} = \frac{2}{4} = \frac{1}{2}$

P(no smokes| no disease) = $\frac{1+1}{2+1(2)} = \frac{2}{4} = \frac{1}{2}$

P(work out|disease) = $\frac{2+1}{4+1(2)} = \frac{3}{6} = \frac{1}{2}$

P(no work out|disease) = $\frac{2+1}{4+1(2)} = \frac{3}{6} = \frac{1}{2}$

P(work out|no disease) = $\frac{2+1}{2+1(2)} = \frac{3}{4}$

P(no work out| no disease) = $\frac{0+1}{2+1(2)} = \frac{1}{4}$

x = (no pain, male, smokes, no workout)

P(disease|x) = P(disease)∗P(no pain|disease)∗P(male|disease)∗P(smokes|disease)∗P(no work out|disease) = $\frac{2}{3} * \frac{1}{3} * \frac{1}{2} * \frac{2}{3} * \frac{1}{2} = \frac{1}{27}$

P(no disease|x) = P(no disease)∗P(no pain|no disease)∗P(male|no disease)∗P(smokes|no disease)∗P(no work out|no disease) = $\frac{1}{3} * \frac{3}{4} * \frac{3}{4} * \frac{1}{2} * \frac{1}{4} = \frac{3}{128}$

Since $\frac{1}{27} > \frac{3}{128}$, the last person is predicted to have the disease

4 Partial Derivatives

Consider the following functions of the variables u, v, and w. Assume the variables x, y, $x^{(i)}$ and $y^{(i)}$ are **constants**: they represent numbers that will not change during the execution of a machine learning algorithm (e.g., the training data). Consider the log as natural logarithm of base e. (2 points for each problem)

$$f(u,v) = 8u^2v^4 + 4v^3 + 6u$$

$$g(u,v,w) = x\log(u) + yuvw^3 + 13x^3$$

$$h(u,v) = \sum_{i=1}^{m} \frac{1}{2}(x^{(i)}u + y^{(i)}v)^2$$

Write the following partial derivatives:

1. $\frac{\partial}{\partial u} f(u,v) = 16uv^4 + 6$

2. $\frac{\partial}{\partial v} f(u, v) = 32u^2v^3 + 12v^2$
3. $\frac{\partial}{\partial u} g(u, v, w) = \frac{x}{u} + yvw^3$
4. $\frac{\partial}{\partial v} g(u, v, w) = yuw^3$
5. $\frac{\partial}{\partial w} g(u, v, w) = 3yuvw^2$
6. $\frac{\partial}{\partial u} h(u, v) = (x^{(1)}u + y^{(1)}v)x^{(1)} + \cdots + (x^{(m)}u + y^{(m)}v)x^{(m)}$
7. $\frac{\partial}{\partial v} h(u, v) = (x^{(1)}u + y^{(1)}v)y^{(1)} + \cdots + (x^{(m)}u + y^{(m)}v)y^{(m)}$

5 Partial Derivative Intuition

Consider the following contour plot of a function $f(u, v)$:

`hw1-contour.pdf`

For each of the following partial derivatives, state whether it is positive, negative, or equal to zero. Briefly explain. These questions can be answered from the contour plot without knowing the formula for the function. (2 points for each problem)

(**Note**: for two numbers a and b we will use the notation $\frac{\partial}{\partial u} f(a, b)$ to mean "the partial derivative of $f(u, v)$ with respect to u at the point where $u = a$ and $v = b$". This notation is succinct but obfuscates the original variable names. A more explicit way to write the same thing is $\frac{\partial}{\partial u} f(u, v)|_{u=a, v=b}$)

1. $\frac{\partial}{\partial u} f(-2, -2)$

   Negative, as moving right in the u-direction leads to a lower contour

2. $\frac{\partial}{\partial v} f(-2, -2)$

   Negative, as moving up in the v-direction leads to a lower contour

3. $\frac{\partial}{\partial u} f(3, -3)$

   Positive, as moving right in the u-direction leads to a higher contour

4. $\frac{\partial}{\partial v} f(3, -3)$

   Negative, as moving up in the v-direction leads to a lower contour

5. To the nearest integer, estimate the values of u and v that minimize $f(u, v)$.

   $u = 1$, $v = 2$

6 Matrix Manipulation I

Matrix multiplication practice (2 points each for the first 5 problems and 5 points for the last problem).

1. Write the result of the following matrix-matrix multiplication. Your answer should be written in terms of u, v, a and b.

$$\begin{bmatrix} 3 & -1 \\ 2 & 5 \\ -2 & 2 \end{bmatrix} \cdot \begin{bmatrix} u & a \\ v & b \end{bmatrix}$$

$$= \begin{bmatrix} 3u - v & 3a - b \\ 2u + 5v & 2a + 5b \\ -2u + 2v & -2a + 2b \end{bmatrix}$$

2. Suppose $A \in \mathbb{R}^{2\times 2}$, $B \in \mathbb{R}^{2\times 4}$. Does the product AB exist? If so, what size is it?

   Yes, AB exists and its size is 2×4

3. Suppose $A \in \mathbb{R}^{3\times 5}$, $B \in \mathbb{R}^{4\times 1}$. Does the product AB exist? If so, what size is it?

   No, AB does not exist

4. Suppose $A \in \mathbb{R}^{3\times 2}$, $y \in \mathbb{R}^3$. Is $y^T A$ a row vector or a column vector?

   Row vector

5. Suppose $A \in \mathbb{R}^{3\times 2}$, $x \in \mathbb{R}^2$. Is Ax a row vector or a column vector?

   Column vector

6. Suppose $(Bx + y)^T A^T = 0$, where A and B are both invertible $n \times n$ matrices, x and y are vectors in $\mathbb{R}^n$, and 0 is a vector of all zeros. Use the properties of multiplication, transpose, and inverse to show that $x = -B^{-1}y$. Show your work.

   Since A and B are both invertible $n \times n$ matrices, then $AB = I$ and $A = B^{-1}$.
   $(Bx + y)^T (B^{-1})^T = 0$
   $(BB^{-1}x + B^{-1}y)^T = 0$
   $(BB^{-1}x)^T + (B^{-1}y)^T = 0$
   $(Ix)^T + (B^{-1}y)^T = 0$
   $x^T + (B^{-1}y)^T = 0$
   $x^T = 0 - (B^{-1}y)^T$
   $(x^T)^T = (-(B^{-1}y)^T)^T$
   $x = -B^{-1}y$

7 Matrix Manipulation II

(10 points) Create a jupyter notebook called **exercise_1.ipynb** and write code to do the following.

1. Enter the following matrices and vectors

$$A = \begin{bmatrix} -2 & -3 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, x = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

2. Compute $C = A^{-1}$

3. Check that $AC = I$ and $CA = I$

4. Compute Ax

5. Compute $A^T A$

6. Compute $Ax - Bx$
7. Compute $||x||$ (use the dot product)
8. Compute $||Ax - Bx||$
9. Print the first column of A (do not use a loop – use array "slicing" instead)
10. Assign the vector x to the first column of B (do not use a loop – use "array slicing" instead)
11. Compute the element-wise product between the first column of A and the second column of A

8 Linear Regression

The problems consider linear regression based on the following hypothesis and function. (10 points each)

$$h_\theta(x) = \theta_0 + \theta_1 x, J(\theta_0, \theta_1) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

Consider the following small data set:

x	y
2	5
-1	-1
1	3

1. Solve for the values of θ_0 and θ_1 that minimize the cost function by substituting the value from the training set into the cost function, setting the derivatives (with respect to both θ_0 and θ_1) equal to zero, and solving the system of two equations for θ_0 and θ_1. Show your work.

   $J(\theta_0, \theta_1) = \frac{1}{2}((h_\theta(2) - 5)^2 + (h_\theta(-1) - (-1))^2 + (h_\theta(1) - 3)^2)$
   $= \frac{1}{2}(((\theta_0 + 2\theta_1) - 5)^2 + ((\theta_0 - \theta_1) + 1)^2 + ((\theta_0 + \theta_1) - 3)^2)$
   $= \frac{1}{2}(((\theta_0)^2 + 4\theta_0\theta_1 - 10\theta_0 + 4(\theta_1)^2 - 20\theta_1 + 25) + ((\theta_0)^2 - 2\theta_0\theta 1 + 2\theta 0 + (\theta_1)^2 - 2\theta_1 + 1) + ((\theta_0)^2 + 2\theta_0\theta 1 - 6\theta 0 + (\theta_1)^2 - 6\theta_1 + 9)$
   $= \frac{1}{2}(3(\theta_0)^2 + 4\theta_0\theta_1 - 14\theta_0 + 6(\theta_1)^2 - 28\theta_1 + 35)$

   $\frac{1}{2}(6\theta_0 + 4\theta_1 - 14) = 0$, $\frac{1}{2}(4\theta_0 + 12\theta_1 - 28) = 0$

   $3\theta_0 + 2\theta_1 - 7 = 0$, $2\theta_0 + 6\theta_1 - 14 = 0$

   $3\theta_0 + 2\theta_1 = 7$, $2\theta_0 + 6\theta_1 = 14 \rightarrow \theta_0 = 7 - 3\theta_1$

   $3\theta_0 + 2\theta_1 = 7 \rightarrow 3(7 - 3\theta_1) + 2\theta_1 = 7 \rightarrow 21 - 9\theta_1 + 2\theta_1 = 7 \rightarrow 14 = 7\theta_1 \rightarrow \theta_1 = 2$

$\theta_0 = 7 - 3(2) = 1$

$\theta_0 = 1, \theta_1 = 2$

2. In this problem you will implement gradient descent for linear regression. Open the notebook **gradient-descent.ipynb** in Jupyter and follow the instructions to complete the problem.

9 Polynomial Regression

(30 points) In this problem you will implement methods for multivariate linear regression and use them to solve a polynomial regression problem. The purpose of this problem is:

1. To practice writing "vectorized" versions of algorithms in Python
2. To understand how feature expansion can be used to fit non-linear hypotheses using linear methods
3. To understand feature normalization and its impact on numerical optimization for machine learning. Open the notebook **multivariate-linear-regression.ipynb** and follow the instructions to complete the problem.