CS178 Fall 2023 midterm

Howard You

TOTAL POINTS

44.5 / 50

QUESTION 1

1 Problem 1.1 **3 / 3**

✓ **- 0 pts** *Correct*

- 1.5 pts In case of ties prefer **earlier** data, prediction for 3 is 1 (not 4)

- 3 pts Wrong

- 3 pts We ask for cross validation error not training error

- 1 pts Algebraic error in calculation

- 1 pts 2^2 + 2^2 + 1 = 9 not 5

- 1.5 pts MSE for x=5 is (4-3)^2=1 not 4

QUESTION 2

2 Problem 1.2 **3 / 3**

✓ **- 0 pts** *Correct*

- 3 pts Wrong

- 1 pts Correct predictions, error in MSE calculation

- 0.5 pts Correct answer, wrong place

- 1.5 pts partially correct predictions / good effort

QUESTION 3

3 Problem 1.3 **3 / 3**

✓ **- 0 pts** *Correct*

- 2 pts partially correct predictions / good effort

- 3 pts Wrong / Missing

- 1 pts It is the same as in (2), but not this one

- 1 pts correct predictions, error in calculation

QUESTION 4

4 Problem 1.4 **3 / 3**

✓ **- 0 pts** *Correct*

- 3 pts Wrong / Missing

- 2 pts Partially correct predictions / some effort

- 1 pts correct predictions, error in calculation

QUESTION 5

5 Problem 1.5 **3 / 3**

✓ **- 0 pts** *Correct*

- 2 pts Wrong because of previous mistakes (but correctly selected smallest mse)

- 3 pts Wrong / Missing

QUESTION 6

6 Problem 2 **7 / 10**

- 1 pts Problem 1 Incorrect

✓ **- 1 pts** *Problem 2 Incorrect*

- 1 pts Problem 3 Incorrect

✓ **- 1 pts** *Problem 4 Incorrect*

- 1 pts Problem 5 Incorrect

✓ **- 1 pts** *Problem 6 Incorrect*

- 1 pts Problem 7 Incorrect

- 1 pts Problem 8 Incorrect

- 1 pts Problem 9 Incorrect

- 1 pts Problem 10 Incorrect

- 0 pts all correct

QUESTION 7

7 Problem 3.1 2 / 2

✓ **- 0 pts** *Correct*

- 0.5 pts Incorrect decision boundary

- 0.75 pts No decision boundary

- 0 pts Decision boundary not in between support vector boundaries

- 0.75 pts Nonlinear decision boundary

- 0.25 pts Missing one support vector

- 1 pts No correct support vectors

- 1.25 pts No answer given for support vectors

- 2 pts Incorrect decision boundary and support vectors

- 0.25 pts One incorrect support vector

- 2 pts Question not answered

QUESTION 8

8 Problem 3.2 4 / 4

✓ **- 0 pts** *Correct*

- 4 pts Answer left blank

- 2.5 pts One coefficient correct

- 1 pts Two coefficients correct

- 3.5 pts Attempted but no coefficients correct

- 1 pts w1 and w2 weights flipped.

- 0.5 pts Incorrect sign for b intercept

- 0.25 pts Correct work but incorrect sign for b intercept

QUESTION 9

9 Problem 3.3 3 / 3

✓ **- 0 pts** *Correct*

- 3 pts Blank Answer

- 2.5 pts Training error is zero fully separable (+0.5 for attempt)

QUESTION 10

10 Problem 3.4 0.5 / 3

- 0 pts Correct

✓ **- 2.5 pts** *LOOXV error is not 0*

- 3 pts Blank answer

- 2.5 pts Incorrect error rate but correct range (within 0-1)

- 3 pts Incorrect answer and outside of possible error rates (error rate between 0 and 1)

- 2.5 pts Answer not given but justification correct

- 1.5 pts Error rate is not 1/6

- 1.5 pts Error rate is not 1/2

- 0.5 pts Simple math error

QUESTION 11

11 Problem 4.1 5 / 5

✓ **- 0 pts** *Correct*

- 0.5 pts 1 error during calculation

- 4 pts Calculated $$P(y|x)$$ instead of $$P(x|y)$$.

QUESTION 12

12 Problem 4.2 4 / 4

✓ **- 0 pts** *Correct*

- 1 pts Mistake carried over from last question

- 1 pts Calculated $$p(y=0|x)$$ instead of $$p(y=1|x)$$

- 1 pts Small mistake, correct formula

- 2 pts Only calculated $$P(y=1,x_1=c,x_2=0)$$

- 2 pts Incorrect evaluation of $$p(x)$$

- 3 pts Incorrect formula

- 4 pts Incorrect, no work

- 4 pts No answer

QUESTION 13

13 Problem 4.3 **4 / 4**

- ✓ **- 0 pts** *Correct*
- **- 2 pts** Calculated $$p(y=0|x_2=b,x_2=1)$$
- **- 4 pts** Incorect
- **- 4 pts** Blank

CS 178 Midterm Exam

Machine Learning and Data Mining: Fall 2023

Wednesday November 8th, 2023

Your name:	**Row/Seat Number:**
Howard You	D108
Your ID #(e.g., 123456789)	**UCINetID (e.g.ucinetid@uci.edu)**
81803650	cyou6@uci.edu

- Please put your name and ID **on every page**.
- Total time is 50 minutes. READ THE EXAM FIRST and organize your time; don't spend too long on any one problem.
- Please **write clearly** and **show all your work**.
- If you need clarification on a problem, please raise your hand and wait for the instructor or TA to come over.
- You may use **one** sheet containing handwritten notes for reference, and a (basic) calculator.
- Turn in your notes and any scratch paper with your exam.

Problems

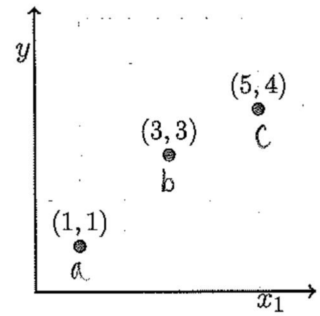
Total, *(50 points.)*

Name: Howard You ID#: 81803650

Problem 1 Cross-Validation, *(15 points.)*

For a regression problem to predict real-valued y given a single real-valued feature x_1, we observe training data (pictured at right):

	x_1	y
a	1.0	1.0
b	3.0	3.0
c	5.0	4.0



(1) Compute the **leave-one-out** cross-validation MSE of a 1-nearest neighbor predictor. (In case of ties, prefer to use the data listed earlier/higher up in the table.) *(3 points.)*

$(1,1): 3 \quad (1-3)^2$ MSE =

$(3,3): 1 \quad (3-1)^2 \quad \frac{1}{3}((1-3)^2+(3-1)^2+(4-3)^2)$

$(5,4): 3 \quad (4-3)^2$

1NN MSE = 3

(2) Compute the **leave-one-out** cross-validation MSE of a 2-nearest neighbor predictor. (In case of ties, prefer to use the data listed earlier/higher up in the table.) *(3 points.)*

$(1,1): \frac{3+4}{2} \quad (1-\frac{7}{2})^2$ MSE =

$(3,3): \frac{1+4}{2} \quad (3-\frac{5}{2})^2 \quad \frac{1}{3}((1-\frac{7}{2})^2+(3-\frac{5}{2})^2+(4-2)^2)$

$(5,4): \frac{1+3}{2} \quad (4-\frac{4}{2})^2$

2NN MSE = $3\frac{1}{2}$

(3) Compute the **leave-one-out** cross-validation MSE of a constant predictor, $f(x) = \theta_0$. *(3 points.)*

$(1,1): \frac{3+4}{2} \quad \frac{7}{2} \quad (1-\frac{7}{2})^2$

$(3,3): \frac{1+4}{2} \quad \frac{5}{2} \quad (3-\frac{5}{2})^2$ MSE $\frac{1}{3}((1-\frac{7}{2})^2+(3-\frac{5}{2})^2+(4-2)^2)$

$(5,4): \frac{1+3}{2} \quad \frac{4}{2} \quad (4-2)^2$

Const MSE = $3\frac{1}{2}$

(4) Compute the **leave-one-out** cross-validation MSE of a linear predictor, $f(x) = \theta_1 x_1 + \theta_0$. *(3 points.)*

MSE =

$(1,1) \quad \frac{4-3}{5-3} = \frac{1}{2} \quad 3 = \frac{1}{2}\cdot 3 + b \quad b = \frac{3}{2} \quad y = \frac{1}{2}x + \frac{3}{2} = \frac{1}{2}\cdot 1 + \frac{3}{2} = 2 \quad \frac{1}{3}((1-2)^2 + (3-\frac{10}{4})^2$

$(3,3) \quad \frac{4-1}{5-1} = \frac{3}{4} \quad 1 = \frac{3}{4} + b \quad b = \frac{1}{4} \quad y = \frac{3}{4}x + \frac{1}{4} = \frac{3}{4}\cdot 3 + \frac{1}{4} = \frac{10}{4} \quad + (4-5)^2)$

$(5,4) \quad \frac{3-1}{3-1} = 1 \quad 1 = 1 + b \quad b = 0 \quad y = x \quad 5$

Linear MSE = $\frac{3}{4}$

(5) If choosing among these models, which would you select (based on cross-validation)? *(3 points.)*

lowest MSE is leave-one-out cross validation MSE of a linear predictor $f(x) = \theta_1 x_1 + \theta_0$

Best model = leave-one-out $f(x) = \theta_1 x_1 + \theta_0$

Name: Howard You ID#: 81803650

Problem 2 True/False, *(10 points.)*

Here, assume that we have m data points $y^{(i)}$, $x^{(i)}$, $i = 1 \ldots m$, each with n features, $x^{(i)} = [x_1^{(i)} \ \ldots \ x_n^{(i)}]$. For each of the scenarios below, circle one of "true" or "false" to indicate whether you agree with the statement.

True or **false**: Using "batch" gradient descent (all m data per step) is less likely to become stuck in local optima than using stochastic gradient descent (1 data point per step).

True or **false**: Optimizing a linear classifier using the hinge loss and L_2 regularization will always converge to a global optimum of the loss.

underfitting = high bias

True or **false**: Using a transformation to increase the number of features available to a linear model will increase its bias.

True or **false**: Using the "perceptron algorithm" to train a linear classifier is always guaranteed to converge.

True or **false**: Given sufficiently many data, a dual-form support vector machine with polynomial kernel $K(x, x') = (1 + x \cdot x'^T)^2$ can approximate any decision function.

True or **false**: The SVM margin optimization problem is an example of a linear program (i.e., a linear optimization objective with linear constraints).

True or **false**: Using a hard-margin SVM instead of a soft-margin SVM will typically reduce the training error.

True or **false**: Dropout can be used to reduce overfitting in neural networks.

True or **false**: Feature selection (electing to ignore certain features of the data in our model) can be used to reduce model variance.

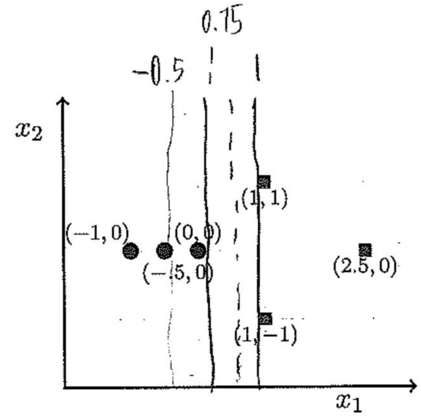
True or **false**: Given sufficiently many layers, a neural network with one hidden node per layer can approximate any function.

Name: Howard You ID#: 81803650

Problem 3 Support Vector Machines, *(12 points.)*

For a classification problem to predict binary y given two real-valued features x_1, x_2, we observe training data (pictured at right):

x_1	x_2	y
0.0	0.0	-1
-0.5	0.0	-1
-1.0	0.0	-1
1.0	-1.0	1
1.0	1.0	1
2.5	0.0	1



Our linear classifier takes the form,

$$f(x; w_1, w_2, b) = \text{sign}(w_1 x_1 + w_2 x_2 + b).$$

(1) Consider the optimal linear SVM classifier for the data, i.e., the one that separates the data and has the largest margin. **Sketch** its decision boundary in the above figure, and **list** the support vectors here. *(2 points.)*

Support vectors: (0,0) (1,1) (1,-1)

(2) Derive the parameter values w_1, w_2, b of this $f(x)$ using these support vectors. *(4 points.)*

$0w_1 + 0w_2 + b = -1$

$1w_1 + 1w_2 + b = 1$

$1w_1 - 1w_2 + b = 1$

$w_1 + w_2 + b = 1$

$w_1 - w_2 + b = 1$

$w_1 + b = 1$

$w_1 - 1 = 1$

$w_1 = 2$

$b = -1$

$w_1 + w_2 + b = 1$

$2 + w_2 - 1 = 1$

$w_2 = 0$

w_1: 2

w_2: 0

b: -1

(3) What is the *training error rate* of a linear SVM on these data? *(3 points.)*

0

Train Err: 0

(4) What is the the *leave-one-out cross validation error rate* for a linear SVM trained on these data? *(3 points.)*

LOOXV Err: 0

0 Since only if you leave out 0,0 -0.5,0 will be the new support vector making the median 0.75 which still classifies 0,0 as -1 correctly

Name: Howard You ID#: 81803656

Problem 4 Naïve Bayes Classifiers, *(13 points.)*

x_1	x_2	y
b	0	0
c	1	0
a	0	0
a	1	0
c	0	1
c	1	1
c	1	1
a	0	1
c	0	1
c	1	1

Consider the table of measured data given at right. We will use the two observed features x_1, x_2 to predict the class y. Feature x_1 can take on one of three values, $x_1 \in \{a, b, c\}$; feature x_2 is binary, $x_2 \in \{0, 1\}$. In the case of a tie, we will prefer to predict class $y = 0$.

(1) Write down the probabilities learned by a naïve Bayes classifier: *(5 points.)*

$p(y=0)$: $\frac{4}{10}$ $p(y=1)$: $\frac{6}{10}$

$p(x_1 = a \mid y = 0)$: $\frac{2}{4}$ $p(x_1 = a \mid y = 1)$: $\frac{1}{6}$

$p(x_1 = b \mid y = 0)$: $\frac{1}{4}$ $p(x_1 = b \mid y = 1)$: $\frac{0}{6}$

$p(x_1 = c \mid y = 0)$: $\frac{1}{4}$ $p(x_1 = c \mid y = 1)$: $\frac{5}{6}$

$p(x_2 = 0 \mid y = 0)$: $\frac{2}{4}$ $p(x_2 = 0 \mid y = 1)$: $\frac{3}{6}$

$p(x_2 = 1 \mid y = 0)$: $\frac{2}{4}$ $p(x_2 = 1 \mid y = 1)$: $\frac{3}{6}$

(2) Using your naïve Bayes model, compute the probability $p(y = 1 \mid x_1 = c, x_2 = 0)$: *(4 points.)*

$$P(y=1 \mid x_1=c, x_2=0) = \frac{P(x_1=c, x_2=0 \mid y=1)\cdot P(y=1)}{P(y=1)\cdot P(x_1=c, x_2=0 \mid y=1) + P(x_1=c, x_2=0 \mid y=0)\cdot P(y=0)}$$

$$= \frac{P(x_1=c \mid y=1)\cdot P(x_2=0 \mid y=1)\cdot P(y=1)}{P(x_1=c \mid y=1)\cdot P(x_2=0 \mid y=1)\cdot P(y=1) + P(x_1=c \mid y=0)\cdot P(x_2=0 \mid y=0)\cdot P(y=0)}$$

$$= \frac{\frac{5}{6}\cdot\frac{3}{6}\cdot\frac{6}{10}}{\frac{5}{6}\cdot\frac{3}{6}\cdot\frac{6}{10} + \frac{1}{4}\cdot\frac{2}{4}\cdot\frac{4}{10}} = \frac{5}{6}$$

$p(y = 1 \mid x_1 = c, x_2 = 0)$: $\frac{5}{6}$

(3) Using your naïve Bayes model, compute the probability $p(y = 1 \mid x_1 = b, x_2 = 1)$: *(4 points.)*

$$P(y=1 \mid x_1=b, x_2=1) = \frac{P(x_1=b, x_2=1 \mid y=1)\cdot P(y=1)}{P(x_1=b, x_2=1 \mid y=1)\cdot P(y=1) + P(x_1=b, x_2=1 \mid y=0) P(y=0)}$$

$$= \frac{P(x_1=b \mid y=1)\cdot P(x_2=1 \mid y=1)\cdot P(y=1)}{P(x_1=b \mid y=1)\cdot P(x_2=1 \mid y=1)\cdot P(y=1) + P(x_1=b \mid y=0)\cdot P(x_2=1 \mid y=0)\cdot P(y=0)}$$

Since $P(x_1=b \mid y=1) = 0$, $P(y=1 \mid x_1=b, x_2=1) = 0$

$p(y = 1 \mid x_1 = b, x_2 = 1)$: 0

This page is intentionally blank, use as you wish.