Midterm exam

Friday Oct 16, 9AM to Sunday Oct. 18 11:59PM (Submission will close automatically by due time. No late submissions. Do not wait until the last minute to submit in case of system glitches.)

Submit your answers as a PDF to Canvas.

The exam is open note and open book. But you cannot talk with anyone about how to solve the problems. Links to precompiled PDFs for the course are on the landing page of the course on Github. You can however, ask the instructor on Canvas or during Zoom sessions (TBD) for clarifications.

Each problem is worth one point, unless otherwise stated. If a problem says that steps must be included, then you get no points if you only provide a final answer.

Pledge

Please put (copy-and-paste, or hand write) this statement on top of your answers. It's better if you add your signature there – but it is not necessary.

"I affirm that the work on this exam is my own and I will not use any people to help me nor will I share any part of this exam or my work with others without permission of the instructor."

Problems

1. According to this class, what are the three types of machine learning? [Hint: Slides for Unit 1]
2. A linear binary classifier's weight vector is $\mathbf{w} = [1, 2, 3]^T$. A sample's augmented feature vector is $\mathbf{x} = [1, 1, 1]^T$. What is the prediction (+1 or −1) for this sample?

   For the linear binary classifier, if $\mathbf{w}^T\mathbf{x} > 0$, the prediction $\hat{y}$ is +1. Otherwise, −1. [Hint: HW2, P.2]
3. Continuing from P. 2 above, if the ground truth label y of the sample $\mathbf{x}$ is −1, what is the error of the classifier on the sample $\mathbf{x}$? Please use the formula on Slide 9 of Unit 2, i.e., $\sum(\mathbf{w}^T\mathbf{x} - y)^2$. Do not use $\sum(\hat{y} - y)^2$.
4. [Bonus] Based on the rule of getting $\hat{y}$ from $\mathbf{w}^T\mathbf{x}$ (last line of P.2), what are the disadvages of using $\sum(\hat{y} - y)^2$ over $\sum(\mathbf{w}^T\mathbf{x} - y)^2$ for error estimation?
5. **[2pt]** Now let's move to decision trees. Suppose we have 6 samples:

Sample Number	feature a	feature b	feature c	class
1	12	3	5	+1
2	5	4	8	+1
3	6	6	7	+1
4	7	5	1	-1
5	8	2	2	-1
6	9	6	3	-1

   Compute the Gini impurity for the four cases:

   - $Pr(class = +1 | b > 5)$
   - $Pr(class = -1 | b > 5)$
   - $Pr(class = +1 | b <= 5)$
   - $Pr(class = -1 | b <= 5)$

   Showing steps is required. [Hint: HW3, P.1,2, and 5]

Half point for each case.

6. Continuing P.5 above, what is $Pr(b > 5)$ and what is $Pr(b <= 5)$? Showing steps is required. Half point for each of the two probabilities. [Hint: HW3, P.3 and P.6]

7. Continuing P.5 and P.6 above, what is the **expectation** of Gini impurity for the feature b and threshold 5? The case $b == 5$ goes with the less than case $b < 5$. [Hint: HW3, P.3 and P.6]

8. Finally, the SVMs. Let's consider the following data to train an SVM.

sample ID (i)	feature a	feature b	feature c	label
1	0.5	0.25	0.125	+1
2	0.4	0.15	0.225	+1
3	0.3	0.75	0.325	-1
4	0.2	0.65	0.425	-1

   If the Lagrangian coefficients (does not necessarily imply they satisfy KKT conditions) for the 4 samples after solving KKT conditions are respectively:

   - $\lambda_1 = 6.13$
   - $\lambda_2 = 0$
   - $\lambda_3 = 4.08$
   - $\lambda_4 = 2.05$

   How many samples are chosen to become support vectors? Justifications are needed to get points. [Hint: Definition of support vectors on Slides 14-15 for Unit 4 SVMs.]

9. Continuing P.8, what are the values of the three elements of the corresponding $\mathbf{w}$?

   Note that $\mathbf{w} = \sum_i \lambda_i y_i \mathbf{x}_i$ where $\mathbf{x}_i$ and y_i are the feature vector and label of the i-th sample above. For example, the feature vector for the 1st sample is $\mathbf{x}_1 = [0.5, 0.25, 0.125]^T$.

   Show your steps. [Hint: HW4, P.1]

10. **[2pt]** Continuing P.8 and P.9, if $w_b = 1.18$, what are the predictions $\mathbf{w}^T\mathbf{x_i} + w_b$ for all samples above? Note that here we simply ask for $\mathbf{w}^T\mathbf{x_i} + w_b$. No need to threshold into $+1$ or -1. Show your steps. Half point for each sample. [Hint: HW4, P.1]

11. **[2pt]** Continuing P.8, P.9 and P.10, according to your answer from P10, does any sample fall into the margin? Show your steps. Half point for checking each sample. [Hint: HW4, P.3]

12. [Bonus] From the result in P.10, do you find anything odd? Explain why it is odd. [Hint: What is the value of the prediction $\mathbf{w}^T\mathbf{x_i} + w_b$ if $\mathbf{x}_i$ is a support vector?]

The END.