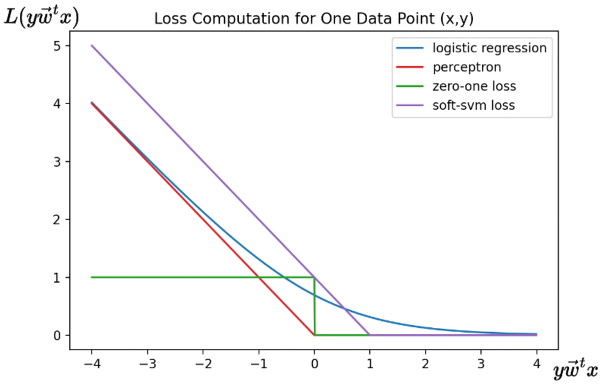
CS461 Section #:	
Name:	
NetID:	

CS461 Quiz Two

0. True / False Questions.

- Both Naive Bayes and logistic regression models learn posterior density $P(C_k|x)$. (**True** / False)
- Perceptron algorithm converges regardless of whether the data is linearly separable. (True / **False**)
- Logistic regression converges regardless of whether the data is linearly separable. (**True** / False)
- When data is linearly separable, logistic regression's sigmoid function becomes infinitely steep around the decision boundary. (**True** / False)
- When an SVM model is trained with a Gaussian kernel, it constructs a maximal margin classifier in an infinite-dimensional feature space. This results in high sensitivity (or high variance) to different training data sets. (True / **False**)

1. The figure shows the loss function: $L(y\vec{w}^t x)$ for the various classification algorithms for one data point (x, y), where $x \in \mathcal{R}^M$ and $y \in \{-1, 1\}$. Please check all correct descriptions.



(1) **The value of $|y\vec{w}^t x|$ implies the distance of x to the decision hyperplane.**

(2) **The positive side of $y\vec{w}^t x > 0$ presents the case of correct classification.**

(3) **The negative side of $y\vec{w}^t x < 0$ presents the case of misclassification.**

(4) Perceptron prefers a larger margin.

(5) Logistic regression does not promote any margin.

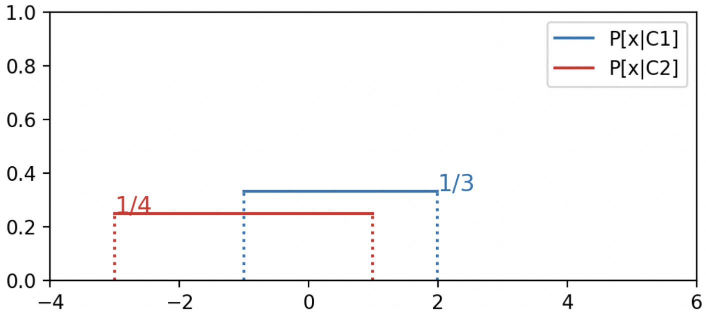
(6) **Perceptron loss considers only misclassified samples.**

(7) Logistic regression loss considers only misclassified samples.

(8) Soft-SVM loss considers the samples on the correct side of the margin.

2. Suppose you classify a sample x using the MAP rule.

> **MAP rule:**
> $$\mathcal{K}^* = \arg\max_k P[C_k|x] \propto P[x|C_k]P[C_k]$$



2.1 Classify the sample $x = 0$ when $P[C_1] = P[C_2] = \frac{1}{2}$. Use the two conditional densities above.

- $\mathcal{K}^* = 1$

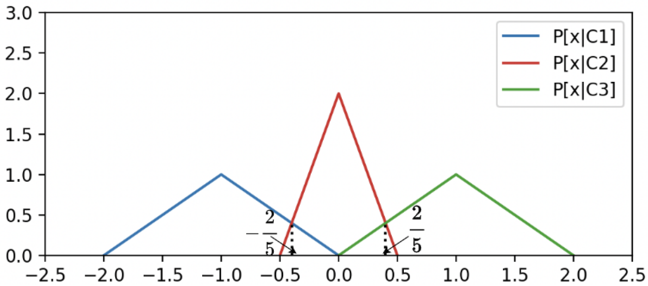
$$P[C1|x=0] = P[x=0|C1]P[C1] = 1/3 * 1/2 = 1/6$$
$$P[C2|x=0] = P[x=0|C2]P[C2] = 1/4 * 1/2 = 1/8$$

2.2 Classify the sample $x = 0$ when $P[C_1] = \frac{2}{3}$ and $P[C_2] = \frac{1}{3}$. Use the same conditional density above.

- $\mathcal{K}^* = 1$

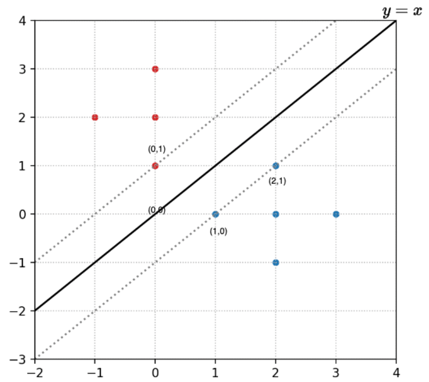
$$P[C1|x=0] = P[x=0|C1]P[C1] = 1/3 * 2/3 = 2/9$$
$$P[C2|x=0] = P[x=0|C2]P[C2] = 1/4 * 1/3 = 1/12$$

2.3 Please define the three decision regions for the classes 1,2, and 3 over the range $-2 \le x \le 2$. Use the conditional densities below. Priors are uniform $P[C_k] = \frac{1}{3}, \forall k$.



- $\mathcal{R}_1 : -2 \le x < -2/5$ (it is okay if $-2 \le x \le -2/5$)
- $\mathcal{R}_2 : -2/5 \le x < 2/5$
- $\mathcal{R}_3 : 2/5 \le x \le 2$

3. Hard margin SVM found the maximum margin decision boundary as below.

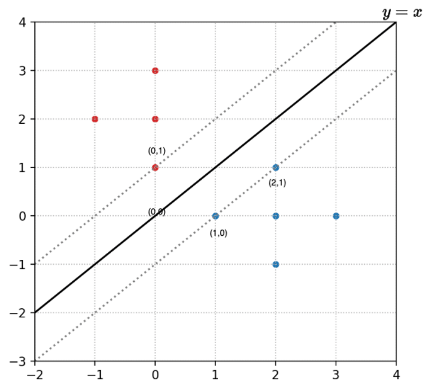


3.1 Please compute the margin of the classifier. You can use the formula: the distance between the hyperplane $w^t x + b = 0$ and x_0: $\dfrac{|w^t x_0 + b|}{||w||}$.

- margin=$\dfrac{1}{\sqrt{2}}$

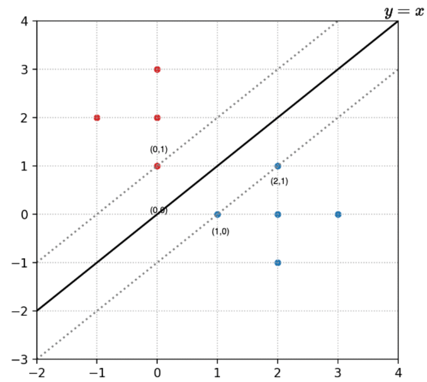
3.2 Please find a minimal data set that will yield the same decision boundary as the same SVM algorithm is used. Circle the data points in the figure below. There are two possible solutions; please provide one example case.

- possible answer: (0,1), (1,0)
- answer: (-1,2), (2,-1)
- answer: (0,2), (2,0)
- answer: (0,3), (3,0)



3.3 Draw a data sample that would produce a different decision boundary in the figure below. The same SVM algorithm will be used. Infinite many solutions are possible; please provide one example case.

sol) If there is any data point within the margin, then the SVM will produce a different decision boundary.



y = x
(0,1)
(0,0)
(1,0)
(2,1)
4
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
4