Week_9_mlt_questions

2024 Mar 24 Quiz 2

1. The training dataset for a binary classification problem is as follows:

$$\{\ (\mathbf{u},\ 1),\quad (-2\mathbf{u},\ 0),\quad (3\mathbf{u},\ 1),\quad (-4\mathbf{u},\ 0)\ \}$$

where $\mathbf{u} \in \mathbb{R}^d$ is a constant, and the labels belong to $0, 1$. Let $\mathbf{w}$ be the weight vector of a linear classifier. What condition should the weight vector satisfy for the zero-one loss to be zero on this dataset?

(a) $\mathbf{w}^T\mathbf{u} < 0$

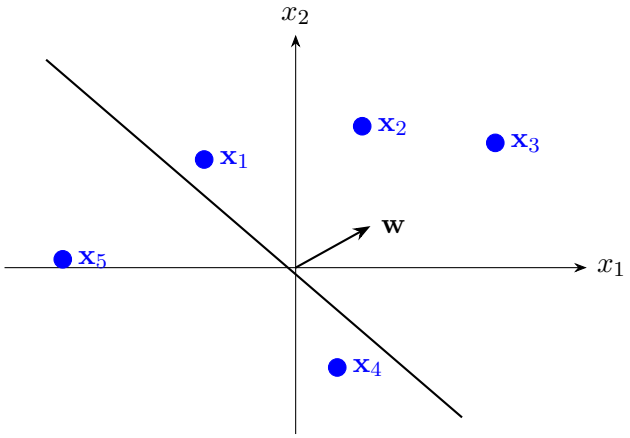
(b) $\mathbf{w}^T\mathbf{u} > 0$

(c) $\mathbf{w}^T\mathbf{u} = 0$

(d) We can never find a $\mathbf{w}$ for which the zero-one loss becomes zero on this dataset.

[Ans: b]

2. **(MSQ)** Consider the following data-points in a binary classification problem. $\mathbf{w}$ is the weight vector corresponding to a linear classifier. The labels are $+1$ and -1.



Which of the following statements are true?

(a) $0 < \mathbf{w}^T\mathbf{x}_1 < \mathbf{w}^T\mathbf{x}_2 < \mathbf{w}^T\mathbf{x}_3$

(b) $\mathbf{w}^T\mathbf{x}_4 = 0$

(c) $0 < \mathbf{w}^T\mathbf{x}_2 < \mathbf{w}^T\mathbf{x}_1 < \mathbf{w}^T\mathbf{x}_3$

(d) $\mathbf{w}^T\mathbf{x}_5 < 0$

(e) $\mathbf{w}^T\mathbf{x}_3 < 0$

[Ans: a, b, d]

2024 Apr 28 ET

3. Consider a logistic regression model for a binary classification problem with two features x_1 and x_2. The feature vector is $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and labels lie in $\{0, 1\}$. The threshold for inference is 0.5. The dummy feature and the weight corresponding to it can be ignored for this problem. Let x_1 be the horizontal axis and x_2 be the vertical axis. You are given two feature vectors:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} -1 \\ \sqrt{3} \end{bmatrix}$$

The weight vector makes an angle of θ with the positive x_1 axis (horizontal). Each θ corresponds to a different classifier. For what range of values of θ are both $\mathbf{x}_1$ and $\mathbf{x}_2$ predicted to belong to class-1?

Hints:

- To draw the weight vector $\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, plot the point (w_1, w_2) and draw an arrow starting at the origin to this point.
- $\tan(60^\circ) = \sqrt{3}$

(a) $30^\circ < \theta < 150^\circ$

(b) $0^\circ < \theta < 60^\circ$

(c) $60^\circ < \theta < 180^\circ$

(d) $0^\circ < \theta < 180^\circ$

(e) $0 < \theta < 360^\circ$

[Ans: a]

4. Consider a linearly separable binary classification dataset with 1000 data points and 100 features. Assume that there exists a w such that $\|w\| = 1$, $y_i(w^T x_i) \geq 0.5 \ \forall i$. Also assume that $\|x\|_2 \leq 2 \ \forall i$. What is the maximum number of mistakes that the Perceptron algorithm can make on this dataset?

[Ans: 16]

2024 Sep 01 ET

5. **(MSQ)** Consider a training dataset of 100 points for a binary classification problem with the following structure:

- The features are in $\mathbb{R}^2$ and the labels are in $\{-1, 1\}$.
- For every data-point $((x_1, x_2), y)$ in the training dataset, $x_1 x_2 > 0$ and $x_1 y > 0$.

Which of the following statements are true?

(a) The dataset is linearly separable with a positive margin.

(b) The perceptron algorithm will terminate after a finite number of iterations when trained on this dataset.

(c) The dataset is linearly separable, but the margin may be zero.

(d) The dataset is not linearly separable.

[Ans: a, b]

6. **(MSQ)** Consider a logistic regression model trained for a binary classification problem with features in $\mathbb{R}^2$ and labels in $\{1, 0\}$. The probability that the test point $\begin{bmatrix}1\\1\end{bmatrix}$ belongs to class 1 is equal to $\frac{1}{1+e^2}$. Which of the following could be the weight vector of the logistic regression classifier? Select all possible answers.

(a) $\begin{bmatrix}-1\\-1\end{bmatrix}$

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

(c) $\begin{bmatrix}1\\1\end{bmatrix}$

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

[Ans: a, b]

7. **Comprehension:**
Consider a linearly separable dataset with a positive margin. The symbol α_i^* in the context of SVMs has its usual meaning. Are the following statements true or false?

i. The weight vector output by the perceptron algorithm on this dataset can be expressed as a linear combination of the data-points where the coefficients of the linear combination are integers.

[Ans: TRUE]

ii. In the case of a hard-margin SVM, if $\alpha_i^* > 0$, the point x_i is a support vector.

[Ans: TRUE]

iii. If a soft-margin SVM is trained on this dataset, the optimal weight vector it returns will be the same as the one returned by a hard-margin SVM, irrespective of the value of the hyperparameter C.

[Ans: FALSE]

2024 Dec 22 ET

8. Consider a logistic regression model trained to detect spam emails. Emails containing harmful phishing links are labeled as spam (1), while regular emails are labeled as not spam (0).

A good spam detector should correctly identify almost all emails with phishing links. Even a single phishing email incorrectly classified as not spam could expose users to significant risks. However, the detector may classify some regular emails as spam. This trade-off is acceptable to ensure safety.

To ensure user safety, if the threshold is set to a low value to prioritize the detection of phishing emails, type 1. If the threshold is set to a high value to minimize false positives (misclassifying regular emails as spam), type 0.

[Ans: 1]

9. While training a perceptron model, the weight vector at some iteration t is $\mathbf{w}^t$. The next two data points picked up by the perceptron algorithm in the course of its execution are $(\mathbf{x}_1, y_1)$ and $(\mathbf{x}_2, y_2)$, where y is the true label:

$$\mathbf{w}^t = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{x}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad y_1 = 1, \quad \mathbf{x}_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \quad y_2 = -1.$$

What are the values of $\mathbf{w}^{t+1}$ and $\mathbf{w}^{t+2}$?

(a) $\mathbf{w}^{t+1} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{w}^{t+2} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$

(b) $\mathbf{w}^{t+1} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{w}^{t+2} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

(c) $\mathbf{w}^{t+1} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{w}^{t+2} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$

(d) $\mathbf{w}^{t+1} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad \mathbf{w}^{t+2} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$

[Ans: a]

2025 Apr 13 ET

10. **(MSQ)** Consider the following data set with three data points:

$$D = \left\{ \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, +1 \right), \left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, -1 \right), \left(\begin{bmatrix} -1 \\ 0 \end{bmatrix}, -1 \right) \right\}.$$

The perceptron algorithm is applied to the dataset with an initial weight vector $w^0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$. While looking for mistakes, cycle through the data points from left to right. Which of the following statements are true?

(a) The algorithm will converge with $w = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$.

(b) The algorithm will converge with $w = \begin{bmatrix} 1 & 2 \end{bmatrix}^T$.

(c) The algorithm will converge with $w = \begin{bmatrix} 0 & 2 \end{bmatrix}^T$.

(d) The algorithm will converge with $w = \begin{bmatrix} 0 & 1 \end{bmatrix}^T$.

(e) The model will predict the label as $+1$, for the new data point $\begin{bmatrix} 2 & -1 \end{bmatrix}^T$.

(f) The model will predict the label as -1, for the new data point $\begin{bmatrix} 2 & -1 \end{bmatrix}^T$.

[Ans: b, e]

11. Consider a logistic regression model that has been trained for a binary classification problem on a dataset in $\mathbb{R}^2$. Given a test data-point as input to the model, it returns 1 as the predicted label if the probability output by the model is greater than 0.55 and 0 otherwise. What will be the predicted label for the test data-point $\begin{bmatrix} 1 & 2 \end{bmatrix}^T$? Assume the weight vector is $\begin{bmatrix} 7 & 4 \end{bmatrix}^T$.

[Ans: 1]

2025 Aug 31 ET

12. Consider the following dataset:

$$D = \left\{ \left(\begin{bmatrix} -2 \\ 1 \end{bmatrix}, +1 \right), \left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, +1 \right), \left(\begin{bmatrix} -1 \\ -1 \end{bmatrix}, -1 \right), \left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, -1 \right) \right\}.$$

The perceptron algorithm is applied on this data set with the weight vector initialized to $\begin{bmatrix} 0 & 0 \end{bmatrix}^T$. Will the algorithm converge after the second round of update of the weight vector? While looking for mistakes, cycle through the data points from left to right.

(a) Yes

(b) No

[Ans: b]

13. **(MSQ)** Consider a dataset $D_1 = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where $x_i \in \mathbb{R}^d$ and is linearly separable by a hyperplane passing through the origin. Construct another dataset $D_2 = \{(2x_1, y_1), (2x_2, y_2), \ldots, (2x_n, y_n)\}$ from D_1. If we run the perceptron algorithm on both the dataset, and the algorithm converges for the points in D_1, then which of the following will hold true?

(a) Perceptron algorithm will converge for D_2.

(b) The perceptron algorithm will give the same classifier when run on D_1 and D_2.

(c) The mistake bound for D_2 will be greater than that of D_1.

[Ans: a, b]

14. Consider a logistic regression model trained to predict whether a customer will buy a product ($Y = 1$) or not ($Y = 0$) based on two features: X_1: number of website visits per week and X_2: time spent on product pages per visit (minutes). The fitted model is:

$$\hat{p} = P(y = 1 \mid x_1, x_2) = \frac{1}{1 + e^{-(0.4x_1 + 0.1x_2)}}.$$

For a customer who spends 4 minutes per visit, find how many weekly visits are required to make the predicted probability of purchase equal to 0.7. Enter the answer correct to 2 decimal places.

[Ans: 1.10 to 1.14]

2025 Dec 18 ET

15. Consider the following training dataset for a binary classification problem:

$$\mathcal{D} = \{(u, 1), (2u, 0), (3u, 1), (4u, 0)\},$$

where $u \in \mathbb{R}^d$ is a fixed nonzero constant and labels belong to $\{0, 1\}$. Let w be the weight vector of a linear classifier. What condition should the weight vector satisfy for the 0-1 loss to be zero on this dataset?

(a) $w^T u < 0$

(b) $w^T u > 0$

(c) $w^T u = 0$

(d) No such weight vector w exists.

[Ans: d]

16. **(MSQ)** Given a training set $\{(x_i, y_i)\}_{i=1}^n$, where $x_i \in \mathbb{R}^d$ and $y_i \in \{0, 1\}$, we want to estimate the parameter w for the logistic regression model. Following is the update rule for estimating w:

$$w^{t+1} = w^t + \eta \left[\sum_{i=1}^{n} x_i \left(y_i - \frac{1}{1 + e^{-w_t^T x_i}} \right) \right].$$

Select the correct options from the following:

(a) If $y_i = 1$, and $\dfrac{1}{1 + e^{-w_t^T x_i}}$ is close to 1, then such points will not contribute much to the update of w.

(b) If $y_i = 1$, and $\dfrac{1}{1 + e^{-w_t^T x_i}}$ is close to 0, then such points will not contribute much to the update of w.

(c) If $y_i = 0$, and $\dfrac{1}{1 + e^{-w_t^T x_i}}$ is close to 1, then such points will not contribute much to the update of w.

(d) If $y_i = 0$, and $\dfrac{1}{1 + e^{-w_t^T x_i}}$ is close to 0, then such points will not contribute much to the update of w.

[Ans: a, d]

17. Consider a binary logistic regression model with the sigmoid function

$$\sigma(z) = \frac{1}{1 + e^{-z}}.$$

Given a single training sample (x, y) where $x = \begin{bmatrix} 1.5 \\ -0.8 \end{bmatrix}$, $y = 1$, and current weight vector $w = \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix}$, compute the cross-entropy loss for this sample given by

$$L(w) = -[y \log_{10}(\hat{y}) + (1 - y) \log_{10}(1 - \hat{y})], \quad \text{where } \hat{y} = \sigma(w^T x).$$

Write your answer correct to three decimal places.

[Ans: 0.273 to 0.279]

18. Consider two datasets $D^{(1)}$ and $D^{(2)}$ where:

$$D^{(1)} = \{(x_1^{(1)}, y_1^{(1)}), \ldots, (x_n^{(1)}, y_n^{(1)})\}, \quad D^{(2)} = \{(x_1^{(2)}, y_1^{(2)}), \ldots, (x_m^{(2)}, y_m^{(2)})\}$$

such that $x_i^{(1)} \in \mathbb{R}^{d_1}$ and $x_i^{(2)} \in \mathbb{R}^{d_2}$, where $d_1 > d_2$ and $n > m$. The maximum number of mistakes made by the perceptron algorithm is higher on dataset $D^{(1)}$ than on dataset $D^{(2)}$.

(a) True

(b) False

[Ans: b]

2026 May 06 ET

19. Consider a dataset with two data-points for a binary classification problem with features in $\mathbb{R}^2$. $(1, 0)$ is a data-point from the positive class while $(-1, 0)$ is from the negative class. We now consider three linear classifiers: perceptron, logistic regression and SVM.
Use gradient ascent to train the logistic regression model with $\eta = 1$. For both perceptron and logistic regression, run exactly one iteration of their respective algorithms starting with $w^{(0)} = (0, 0)$ in each case. That is, you should stop after you get $w^{(1)}$. Train the SVM in the usual way.
Use 0 as the negative class label for logistic regression and -1 for SVM and perceptron. Use the following rule for determining the predicted labels for the three classifiers:
If $w^* \in \mathbb{R}^2$ is the weight vector with the largest margin, which of the following is true?

SVM and Perceptron	**Logistic Regression**
$\hat{y} = \begin{cases} 1, & w^T x \geq 0 \\ -1, & w^T x < 0 \end{cases}$	$\hat{y} = \begin{cases} 1, & w^T x \geq 0 \\ 0, & w^T x < 0 \end{cases}$

(a) All three classifiers output w^*

(b) Only SVM outputs w^*

(c) Only SVM and perceptron output w^*

(d) Only perceptron and logistic regression output w^*

[Ans: a]

20. **(MSQ)** Consider a dataset in $\mathbb{R}^2$ defined as:

$$D = \{([1, 1]^T, +1), ([2, 1]^T, +1), ([1, -1]^T, -1), ([2, -1]^T, -1)\}.$$

Which of the following weight vectors $w \in \mathbb{R}^2$ can correctly separate the two classes using a linear decision boundary? (Select all that apply.)

(a) $w = [0, 1]^T$

(b) $w = [1, 0]^T$

(c) $w = [1, 5]^T$

(d) $w = [-1, 1]^T$

[Ans: a, c]

21. **(MSQ)** Consider a binary classification dataset in $\mathbb{R}^4$ that is linearly separable. Suppose there exist two distinct weight vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ that both correctly classify all training points with a positive margin. The Perceptron algorithm is applied to this dataset. Which of the following statements are correct? (Select all that apply.)

(a) The Perceptron algorithm is guaranteed to converge to $\mathbf{w}_1$.

(b) The final weight vector obtained by the Perceptron algorithm depends on the initialization and the order in which the data points are presented.

(c) The Perceptron algorithm is guaranteed to converge to some separating hyperplane in a finite number of updates.

(d) The Perceptron algorithm always converges to the maximum margin separator among all possible separating hyperplanes.

[Ans: b, c]

22. Consider a binary classification problem using logistic regression. The model predicts the probability of class +1 as:

$$P(y = +1 \mid x) = \frac{1}{1 + e^{-w^T x}}$$

A threshold of 0.7 is used for classification. The predicted class label is given by:

$$P(y \mid x) = \begin{cases} +1, & \text{if } P(y = +1 \mid x) \geq 0.7 \\ -1, & \text{otherwise} \end{cases}$$

The values of $w^T x$ for the following datapoints are: $x_1 : -1.2,\ x_2 : 0.0,\ x_3 : 2.5,\ x_4 : 0.4,\ x_5 : 0.8$. How many datapoints will be classified as class +1?

[Ans: 1]

2026 Sep 13 ET

23. **(MSQ)** Let $S = (x_1, y_1), \ldots, (x_n, y_n)$ be n points in $\mathbb{R}^d$ that are linearly separable by a hyperplane through the origin. Construct $S' = (cx_1, y_1), \ldots, (cx_n, y_n)$ where $c > 1$ is a constant. The perceptron algorithm (initialised at $w = 0$, learning rate 1) is run on both S and S', using the same ordering of points in both runs. Which of the following statements are true?

(a) The mistake bound of the perceptron on S' is strictly larger than its mistake bound on S.

(b) S' may fail to be linearly separable.

(c) The perceptron is guaranteed to converge on S'.

(d) The perceptron produces exactly the same decision boundary on S and on S'.

[Ans: c, d]

24. **(MSQ)** Let

$$g(z) = \frac{1}{1 + e^{-z}}$$

be the sigmoid (logistic) function. Select all statements below that are true for every $z \in \mathbb{R}$.

(a) $g(-z) = 1 - g(z)$.

(b) g is a non-decreasing function but not strictly increasing.

(c) g is a strictly increasing function.

(d) g is convex on all of $\mathbb{R}$.

[Ans: a, c]

25. In logistic regression, a point is classified positive only if $P(y = 1 \mid x) = \sigma(w^T x) \geq 0.75$. What score threshold θ (with $w^T x = \theta$ at the boundary of this rule) does this correspond to? (Round to 2 decimal places.)

[Ans: 1.08 to 1.10]

26. Consider a binary classification dataset in which the two classes are perfectly linearly separable. You are considering a Perceptron, Logistic Regression, and a Decision Tree. If the objective is to learn a linear decision boundary and no probability estimates are required, which of the following classifiers are appropriate choices?

(a) Perceptron only

(b) Logistic Regression only

(c) Both Perceptron and Logistic Regression

(d) Decision Tree only

[Ans: c]