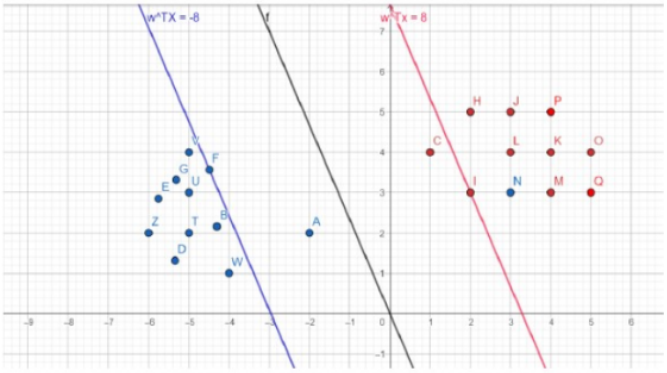
Week_11_mlt_questions

2024 Apr 28 ET

1. **(MSQ)** Consider the following dataset on which the soft margin SVM is applied.



Which of the following statements is/are true about this dataset?

(a) Points $\{F, A, C, I\}$ are the only support vectors.

(b) Points $\{A, N\}$ are a subset of support vectors.

(c) Points $\{F, A, N\}$ are a subset of support vectors.

(d) Points except $\{F, A, C, I, N\}$ do not play any role in determining optimal weightvector.

(e) Points except $\{F, A, C, I\}$ do not play any role in determining optimal weightvector.

[Ans: b, c, d]

2. **(MSQ)** Consider a soft-margin Support Vector Machine (SVM) for a binary classification problem with a dataset in a two-dimensional space. The optimization problem for the soft-margin SVM is formulated as:

$$\text{Minimize } \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

subject to the constraints:

$$y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0 \text{ for all } i$$

where C is a positive constant. Let w^*, ξ^* be the optimal solutions, and α^*, β^* be the optimal dual solutions of the soft-margin SVM problem.

Which of the following statements about the soft-margin SVM is **correct**?

(a) If i^{th} data point lies on one of the supporting hyperplanes, then $\alpha_i^* = 0$.

(b) If i^{th} data point lies on the correct supporting hyperplane, it does **not** pay any bribes.

(c) A smaller value of C allows for a larger margin, potentially leading to less misclassifications on the training data.

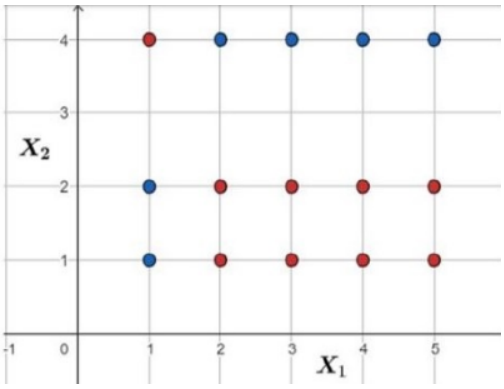
(d) For a dataset with n data-points, there are $2n$ constraints for soft-margin SVM.

(e) As C approaches ∞ the soft-margin SVM is equal to the hard-margin SVM.

(f) C can be negative, as long as the bribe (ξ) each data point pays is non-negative.

[Ans: b]

3. **Comprehension:**
Consider the following two-dimensional dataset with two classes: $+1$ for blue points and -1 for red points. An AdaBoost algorithm was run on this dataset using decision stumps as weak learners.



When training the new weak learner $h_t(x)$ (decision stump at t^{th} iteration), we choose the split that minimizes the weighted miss-classification error with respect to current weights D_t i.e. choose h_t that minimizes $\sum_{i=1}^{n} D_t(i)\mathbf{1}(h_t(x_i) \neq y_i)$. Based on the above data, answer the below given question.

(i) Consider the following two-dimensional dataset with two classes: $+1$ for blue points and -1 for red points. An AdaBoost algorithm was run on this dataset using decision stumps as weak learners. To train the second decision stump, which pair of points will be assigned equal weights to create the training data-set?

(a) $[2, 2]^T$, $[1, 2]^T$

(b) $[2, 2]^T$, $[1, 4]^T$

(c) $[1, 1]^T$, $[1, 4]^T$

(d) $[3, 1]^T$, $[3, 4]^T$

[Ans: a]

(ii) Consider a single iteration of the AdaBoost algorithm that was run on three sample points, starting with uniform weights on the sample points. The labels are either $+1$ or -1. In the table below, some values have been omitted.

Data point	**True label**	**Predicted label**	**Initial weight**	**Updated weight**
x_1	?	1	$\frac{1}{3}$	$\frac{1}{2}$
x_2	-1	-1	$\frac{1}{3}$	?
x_3	-1	?	$\frac{1}{3}$	$\frac{1}{4}$

Based on the above data, what will be the updated weight for point x_2? Enter your answer correct to two decimal places.

[Ans: 0.25]

2024 Sep 01 ET

4. **(MCQ)** Match the characteristics given below with the appropriate ensemble technique.

   1. Decision stumps
   2. Deep decision trees
   3. Parallel execution
   4. Sequential execution

(a) Bagging → (2), (3); Boosting → (1), (4)

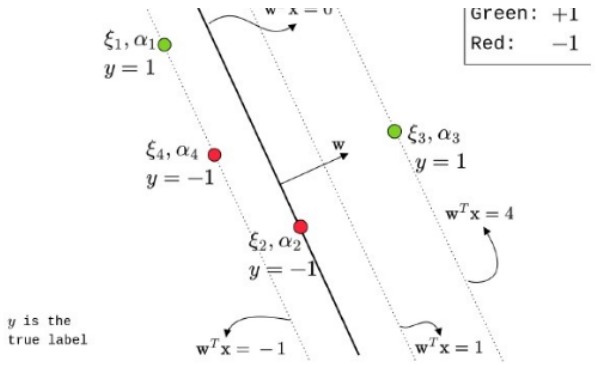
(b) Bagging → (1), (4); Boosting → (2), (3)

(c) Bagging → (1), (3); Boosting → (2), (4)

(d) Bagging → (2), (4); Boosting → (1), (3)

[Ans: a]

5. **Comprehension:**
Consider a soft-margin SVM with $C = 3$ that has been trained on a dataset with features in $\mathbb{R}^2$. The decision boundary and the supporting hyperplanes are displayed below. Four points from the training dataset are also displayed. Green data-points belong to class 1 and red data-points belong to class -1. Symbols ξ and α have their usual meanings. Assume that $\mathbf{w}, \xi_i, \alpha_i$ represent the optimal values.



Based on the above data, answer the given subquestions.

(i) What is ξ_1? [Ans: 2]

(ii) What is ξ_2? [Ans: 1]

(iii) What is ξ_3? [Ans: 0]

(iv) What is α_1? If it cannot be determined exactly, enter -1. [Ans: 3]

(v) What is α_3? If it cannot be determined exactly, enter -1. [Ans: 0]

(vi) What is α_4? If it cannot be determined exactly, enter -1. [Ans: -1]

2024 Dec 22 ET

6. **(MCQ)** In each round of AdaBoost, the weight for a particular training observation is decreased from round t to round $t + 1$ if the observation was:

(a) classified incorrectly by the weak learner trained in the round.

(b) classified correctly by the weak learner trained in the round.

(c) classified incorrectly by a majority of the weak learners trained up to the round.

(d) classified correctly by a majority of the weak learners trained up to the round.

[Ans: b]

7. **(MSQ)** Which of the following options are correct? Select all that apply.

(a) Underfitted models generally have less bias and high variance.

(b) A decision tree with the maximum possible length may lead to overfitting.

(c) Weak learners are those whose performance is slightly better than the random classifier.

(d) All the estimators in bagging can be trained parallelly.

[Ans: b, c, d]

8. **Comprehension:**
Consider a soft-margin SVM that has been trained on a dataset in $\mathbb{R}^2$. A subset of the data points and the decision boundary (solid line) is shown below:

For each point, consider ξ_i to represent the slack variable or the "minimum bribe" required to move the point onto the correct supporting hyperplane. Answer the given sub-questions based on the figure:

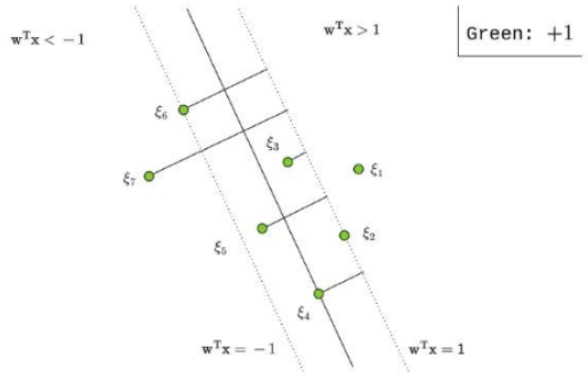


Figure 1: Decision boundary and supporting hyperplanes.
Green points belong to class +1.

(i) What is the value of ξ_4? [Ans: 1]

(ii) What is the value of ξ_1? [Ans: 0]

(iii) What is the value of ξ_2? [Ans: 0]

(iv) What is the value of ξ_6? [Ans: 2]

(v) Which of the following is correct? **(MSQ)**

(a) $\xi_7 > 2$
(b) $1 < \xi_5 < 2$
(c) $\xi_7 < 0$
(d) $1 < \xi_7 < 2$

[Ans: a, b]

2025 Apr 13 ET (QIM4)

9. **(MSQ)** Let w^*, ξ^* be the optimal primal solutions, and α^*, β^* be the optimal dual solutions of the soft-margin SVM problem. Select the options which are **always true**.

(a) If $\alpha_i^* = C$, the i^{th} point is incorrectly classified by w^*.

(b) If $\alpha_i^* = 0$, the i^{th} point is correctly classified by w^*.

(c) If the i^{th} data point lies on one of the supporting hyperplanes, then $\alpha_i^* = 0$.

(d) If the i^{th} data point lies on one of the supporting hyperplanes, it does not pay any bribes.

(e) If the i^{th} data point lies between the supporting hyperplanes, it always pays a positive bribe.

[Ans: b, d, e]

10. **(MCQ)** Weak learners used in the ensemble method typically perform slightly better than random guessing.

(a) TRUE

(b) FALSE

[Ans: a (TRUE)]

11. **Comprehension:**
Consider a single iteration of the AdaBoost algorithm run on five sample points $\{x_1, x_2, x_3, x_4, x_5\}$ with labels 1 or -1, starting with uniform weights. Two classifiers h_1 and h_2 predict the labels as shown in the below table:

Data point	**True label**	**Predicted label(h_1)**	**Predicted label(h_2)**
x_1	1	-1	1
x_2	-1	1	1
x_3	-1	-1	1
x_4	1	1	-1
x_5	-1	-1	1

Based on the above data, answer the given subquestions.

(i) (MCQ) Which classifier has the lower training error?

(a) h_1

(b) h_2

[Ans: a (h_1)]

(ii) (MCQ) The classifier with the lower training error rate is used in the next round of the AdaBoost algorithm. Determine the classifier h^* obtained at the end of this round.

(a) $h^*(x) = sign(0.202 * h_2)$

(b) $h^*(x) = sign(0.202 * h_1)$

(c) $h^*(x) = sign(-0.69 * h_1)$

(d) $h^*(x) = sign(-0.69 * h_2)$

[Ans: b]

2025 Aug 31 ET (QDF4)

12. **(MSQ)** Consider a 1D dataset which consists of 2 positive data points (points with label 1) $\{1, 2\}$ and 3 negative data points (points with label -1) $\{-3, -2, -1\}$. Suppose that we want to learn a soft-margin linear SVM for this data set. Consider the formulation of the soft-margin SVM:

$$\min_{w,\epsilon} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \epsilon_i, \qquad C \geq 0$$

$$\text{subject to } (w^T x_i)y_i + \epsilon_i \geq 1, \quad \forall i$$

$$\epsilon_i \geq 0, \quad \forall i.$$

Select the correct options from the following:

(a) Number of support vectors for $C = 0$ is 0.

(b) Number of support vectors for $C = 0$ is 2.

(c) Number of support vectors for $C = \infty$ is 5.

(d) Number of support vectors for $C = \infty$ is 2.

[Ans: a, d]

13. **(SA)** A random forest model consists of three decision trees. For a given test point, each tree outputs the following class probabilities for class $+1$ and class -1:

$$\begin{aligned}
&\text{Tree 1:} \quad P(+1) = 0.7, \quad P(-1) = 0.3 \\
&\text{Tree 2:} \quad P(+1) = 0.4, \quad P(-1) = 0.6 \\
&\text{Tree 3:} \quad P(+1) = 0.8, \quad P(-1) = 0.2
\end{aligned}$$

The random forest predicts the final label by averaging the class probabilities across the trees and selecting the class with the higher average probability. What is the final predicted label for the test point? (Give the answer as 1 for class $+1$ or -1 for class -1.)

[Ans: 1]

2025 Dec 18 ET

14. **(MSQ)** In a random forest classifier with p total features, each split is chosen by evaluating a randomly selected subset of m features. Consider how changing m affects the behavior of the forest. Which of the following statements are correct?

(a) Increasing m reduces the correlation between any two trees in the forest.

(b) Decreasing m reduces the correlation between any two trees in the forest.

(c) If m is too small, individual trees may underfit the data.

(d) Increasing m always decreases the variance of the random forest model.

(e) Decreasing m makes trees more similar to each other.

[Ans: b, c]

15. **Comprehension:**
Consider the following one-dimensional dataset with binary labels $y \in \{+1, -1\}$.

Points	x	y
1	1	$+1$
2	2	$+1$
3	3	-1
4	4	-1
5	5	$+1$

AdaBoost uses decision stumps as weak learners. Assume the initial distribution of the dataset assigns equal weights to all data points, i.e.,

$$D_0(i) = \frac{1}{5}, \quad \text{for } i = 1, 2, 3, 4, 5.$$

At iteration $t = 1$, AdaBoost chooses the decision stump

$$h_1(x) = \begin{cases} +1, & x < 3, \\ -1, & x \geq 3. \end{cases}$$

Based on the above data, answer the given subquestions.

(i) (MCQ) Which point will be misclassified by the first decision stump?

(a) Points 4

(b) Points 5

(c) Points 3

(d) Points 1

[Ans: b]

(ii) (MSQ) After AdaBoost updates the sample weights for iteration $t = 2$, which pair of points will have the smallest (and equal) weights while training the second stump h_2?

(a) Points 1 and 2

(b) Points 3 and 4

(c) Points 3 and 5

(d) Points 1 and 5

[Ans: a, b]

2026 May 06 ET

16. **(MCQ)** Which of the following statements is correct?

(a) Bagging reduces variance, Boosting reduces bias

(b) Bagging reduces bias, Boosting reduces variance

(c) Both reduce only variance

(d) Both reduce only bias

[Ans: a]

2026 Sep 13 ET

17. **(MCQ)** For an optimal soft-margin SVM solution (no bias term), a training point satisfies $0 < \alpha_i^* < C$. Here ξ_i denotes the slack variable for this point, and β_i is the Lagrange multiplier on the constraint $\xi_i \geq 0$. Which statement is necessarily true?

(a) The point is misclassified by w^*.

(b) $\xi_i^* > 0$.

(c) $y_i({w^*}^\top x_i) = 1$ and $\xi_i^* = 0$.

(d) $\beta_i^* = 0$.

[Ans: c]

18. **(SA)** *Note: coordinates of x_1 vector are intentionally not provided.*

Consider a soft-margin SVM for binary classification in $\mathbb{R}^3$ with **no bias term** and regularisation parameter $C = 10$. The training set has three points:

$$x_1, (y_1 = +1), \quad x_2 = (0.2, 0.3, 0.4), (y_2 = +1), \quad x_3 = (0.15, 0.28, 0.35), (y_3 = -1).$$

At the optimal solution w^* the margin positions are

$$\begin{aligned} y_1({w^*}^\top x_1) &= 1.5, \\ y_2({w^*}^\top x_2) &= 0.36, \\ y_3({w^*}^\top x_3) &= -0.306. \end{aligned}$$

Compute the perpendicular distance between the two margin hyperplanes $w^\top x = +1$ and $w^\top x = -1$. *(Round to 2 decimal places.)*

[Ans: 2.71 to 2.73]

19. **(MSQ)** Which of the following statements about bagging are correct?

(a) Bagging reduces the variance of the ensemble's prediction.

(b) Bagging substantially reduces the bias of the base learner.

(c) Bagging tends to help most when the base learners are diverse (weakly correlated with one another).

(d) Bagging tends to help most with high-variance, low-bias base learners such as deep, unpruned decision trees.

(e) Averaging all base learner reduces variance, if every base learner produces identical predictions.

[Ans: a, c, d]