CS-233 Theoretical Exercise

May 2024

1 Short questions

Question 1: Are the following statements true?

1. AdaBoost is a type of ensemble learning algorithm that combines multiple weak classifiers to create a strong classifier.
2. In AdaBoost, all weak classifiers contribute equally to the final model.
3. The weight of each weak classifier in AdaBoost is determined based on its weighted classification error.
4. The weight of each weak classifier in AdaBoost is determined based on its unweighted classification error.
5. AdaBoost can perfectly classify any dataset if given enough iterations.

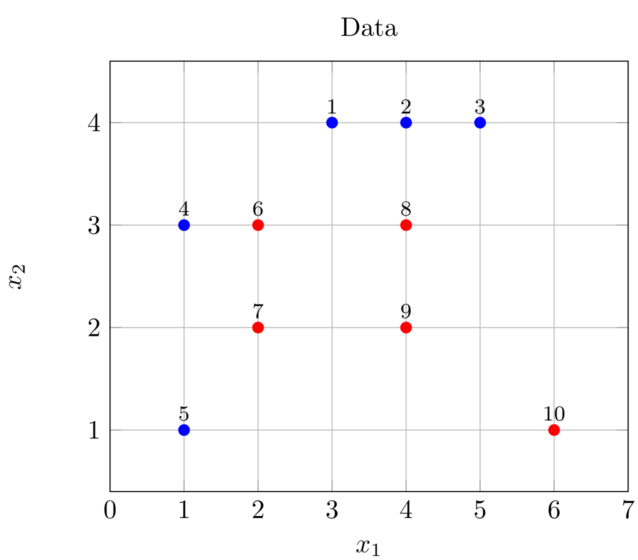
Question 2: AdaBoost is known to be sensitive to outliers. Why does this happen?

Question 3: Suppose that in your implementation of AdaBoost, you make the following mistake: you replace "$\neq$" with "$=$" in the weight update:

$$w_n^{t+1} = w_n^t \exp(\alpha_t I(t_n = y_t(x_n)))$$

What would happen then? How would the weak classifiers look like?

2 AdaBoost: simple example



Consider the dataset presented in the above figure, where each data point has two coordinates. Blue dots correspond to $t = 1$, and red dots correspond to $t = -1$. All your weights are initialized equally, i.e. $w_n^1 = 1$

for all n. Suppose that your weak classifier can only be of the form

$$y_t(\boldsymbol{x}) = \text{sign}(x_i - c) \text{ or } y_t(\boldsymbol{x}) = \text{sign}(c - x_i)$$

where $i \in \{1, 2\}$ and $c \in \mathbb{R}$. In other words, the decision boundary of each weak classifier is either vertical (e.g. $x_1 = c$) or horizontal (e.g. $x_2 = c$).

Question 1: What is the first weak classifier y_1 selected by the AdaBoost algorithm? Evaluate the associated ϵ_1 and α_1.

Question 2: Which of the weights are updated during this step? What are the updated weights w_n^2 for these?

Question 3: What is the second weak classifier y_2 selected by the algorithm? Evaluate the associated ϵ_2 and α_2. What are the weights w_n^3 for all n after the update?

Question 4: What is the third weak classifier y_3 selected by the algorithm? Evaluate the associated ϵ_3 and α_3.

Question 5: Express the final classifier after three steps. Show that all points in the dataset are correctly classified by the final classifier.
Hint: observe that $\alpha_i + \alpha_j > \alpha_k$ for any $i, j, k \in \{1, 2, 3\}$.

3 Support Vector Machines

Consider a set of two-dimensional training examples $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^2$ and $y_i \in \{-1, +1\}$. Assume that the data is linearly separable and that we have scaled the problem so that for the points closest to the decision boundary (the support vectors), we have for $i \in \{1, \dots, n\}$:

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) = 1.$$

3.1 Margin Calculation

Question 1. Show that the perpendicular distance from any point $\mathbf{x}$ to the hyperplane defined by $\{\mathbf{v} \in \mathbb{R}^2 \colon \mathbf{w}^\top \mathbf{v} + b = 0\}$ is given by

$$\text{Distance} = \frac{|\mathbf{w}^\top \mathbf{x} + b|}{\|\mathbf{w}\|}.$$

Question 2. Using the scaling condition for the support vectors, derive the expression for the margin m, the distance from the decision boundary to the support vectors.

3.2 Optimal Hyperplane

Question 3. Explain why maximizing the margin is equivalent to minimizing $\|\mathbf{w}\|$ (or, for convenience, minimizing $\frac{1}{2}\|\mathbf{w}\|^2$).
Question 4. Write down the optimization problem for the hard-margin SVM based on this insight.

3.3 Interpreting the Constraints

Question 5. Describe in plain language what the constraint

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$$

means for each training example.
Question 6. What would be the practical consequence if a training point does not satisfy this constraint?

3.4 Bonus Question: Soft-Margin SVM and the Role of Parameter C

In real-world applications, datasets are often not perfectly linearly separable. To handle this, the soft-margin SVM introduces slack variables $\xi_i \geq 0$ to allow some points to violate the margin condition. The modified constraints are:

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \forall\, i.$$

Additionally, a penalty term is added to the objective function.

Question 7. Write down the modified optimization problem (primal formulation) for the soft-margin SVM, including the slack variables and the regularization parameter C.
Question 8. Explain the meaning of the slack variables ξ_i. Discuss how the parameter C influences the trade-off between maximizing the margin and minimizing classification errors. Describe what happens to the model behavior as C increases or decreases.