Assignment – Machine Learning

Deadline: Dec 7, before 23:59 PM

Submission: answer the following questions.

Exercise 1 [1 point]:

Consider a binary classification problem on data set $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^N$ with raw inputs $\boldsymbol{x}_i \in \mathbb{R}^D$ and outputs $y_i \in \{0, 1\}$. You decide to transform your raw features to get a new dataset $\{(\boldsymbol{\phi}(\boldsymbol{x}_i), y_i)\}_{i=1}^N$. Specifically, you choose $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ for some $\boldsymbol{A} \in \mathbb{R}^{M \times D}$ with dimension $M > D$.

If the original dataset is *not* linearly separable, might the transformed dataset be linearly separable? YES/NO then explain (1 sentence).

Exercise 2 [4 points]:

Consider a 3-class SVM trained with one-vs-rest strategy. Suppose that the linear descriminant for each class $k \in \{1, 2, 3\}$ is defined, respectively, by:

$$\begin{aligned} \boldsymbol{w}_1 &= (2, \tfrac{1}{2}) \qquad & b_1 &= 2 \\ \boldsymbol{w}_2 &= (3, -3) \qquad & b_2 &= \tfrac{9}{2} \\ \boldsymbol{w}_3 &= (-1, -\tfrac{3}{2}) \qquad & b_3 &= -2 \end{aligned}$$

a. [3 marks] What values do the three linear discriminants y_1, y_2, y_3 take for input $\boldsymbol{x} = (-\frac{1}{2}, -2)$? Show your calculations for all three numbers.

b. [1 mark] What class would the combined model predict? Explain.

Exercise 3 [4 points]:

a. [1 mark] Random forests are bagged decision trees. TRUE or FALSE?

b. [1 mark] Compared to training a single model on all the training data, does bagging tend to reduce the *training error*, the *testing error*, or *both*?

c. [2 marks] Consider training set:

```
>>> X = [[0,  0],
         [1, 10],
         [2, 20],
         [3, 30]]
>>> y = [0, 1, 2, 3]
```

Suppose we draw a bootstrap sample of $(\boldsymbol{X}, \boldsymbol{y})$, giving:

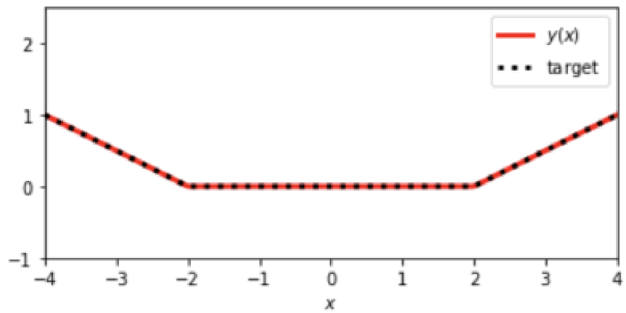
```
>>> X_bootstrap
[[3, 30],
 [1, 10],
 [3, 30],
 [0,  0]]
>>> y_bootstrap
[3, 1, 3, 0]
```

What *sample_weight*, when applied to the original training set, would be equivalent to training on this bootstrap sample? i.e., give 4 numbers:

```
estimator.fit(X_bootstrap, y_bootstrap)          # bootstrap
estimator.fit(X, y, sample_weight=[?, ?, ?, ?])  # same
```

Exercise 4 [6 points]:

Part (c) pertains to the target function below, and the neural network $y(x)$ that manages to approximate it (in this case, exactly).



a. [1 mark] What is the specific role of backpropagation? Brief sentence.

b. [2 marks] Suppose input $\boldsymbol{x}$ has 50 features and you are doing 10-way classification. How many parameters would a 50-100-20-10 sigmoid network contain? Show calculations (hidden layer 1 + hidden layer 2 + output layer).

c. [3 marks] Give specific values for the parameters of the 1-2-1 ReLU neural network that matches the $y(x)$ from the figure. The architecture is:

$$\begin{aligned} z_1 &= \mathrm{ReLU}(w_1 x + b_1) \\ z_2 &= \mathrm{ReLU}(w_2 x + b_2) \\ y &= w_3 z_1 + w_4 z_2 + b_3 \end{aligned}$$

You can type `w1 = ...` to indicate the value of w_1, etc.

Exercise 5 [5 points]:

Part (c) is about the function below:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{32}x^2} + \frac{3}{4\sqrt{2\pi}} e^{-\frac{1}{2}x^2+4x-4}$$

(Here symbol $\pi \approx 3.14$ is a constant, not a component mixture weight.)

a. [1 mark] Are Gaussian mixture models (GMMs) *parametric* or *non-parametric*?

b. [1 mark] In the EM algorithm, how are the $\boldsymbol{\mu}_k$ typically initialized in practice?

c. [3 marks] Is $f(x)$ a valid Gaussian mixture density? YES/NO then

- If YES, give the six parameters (w_1, μ_1, σ_1) and (w_2, μ_2, σ_2) of the model, where w is mixture weight, μ is mean, and σ is standard deviation.
- If NO, explain why it's not a Gaussian mixture density (1 sentence).