ENGN 2520 Pattern Recognition and Machine Learning

Exam

Zhuo Wang

Problem 1

(a) What is the difference between the training error and test error?

Training error is the error computed using the model's training set of data. This is the error that the model makes using the "seen" data from its training. It assesses how well the training set of data matches the model.

Test error is the error determined using a different dataset that the model did not see during training. We named it the test dataset. This helps in assessing how effectively the model generalizes to unobserved data.

The test error shows how effectively the model generalizes to new, unknown data, whereas the training error shows how well the model fits the training data.

(b) What is overfitting? How can we avoid it in practice?

Overfitting refers to the phenomenon that the model performs well on the training set but poorly on the test set, also known as "high variance". This often occurs when the model overlearns the training data to the point that it treats the noise and outliers in the training data as useful information. As a result, the model's capacity to generalize and generate accurate predictions about new, unseen data is diminished.

We can avoid overfitting by increasing the training data. Because more training data can provide more comprehensive information about the sample distribution, it helps the model learn more general laws. We can also add regular terms to avoid overfitting. Add a regular term to the objective function so that the coefficients do not reach large values. It can help prevent the model from over-fitting the training data, thereby improving the model's ability to generalize.

(c) What is the margin in a support vector machine?

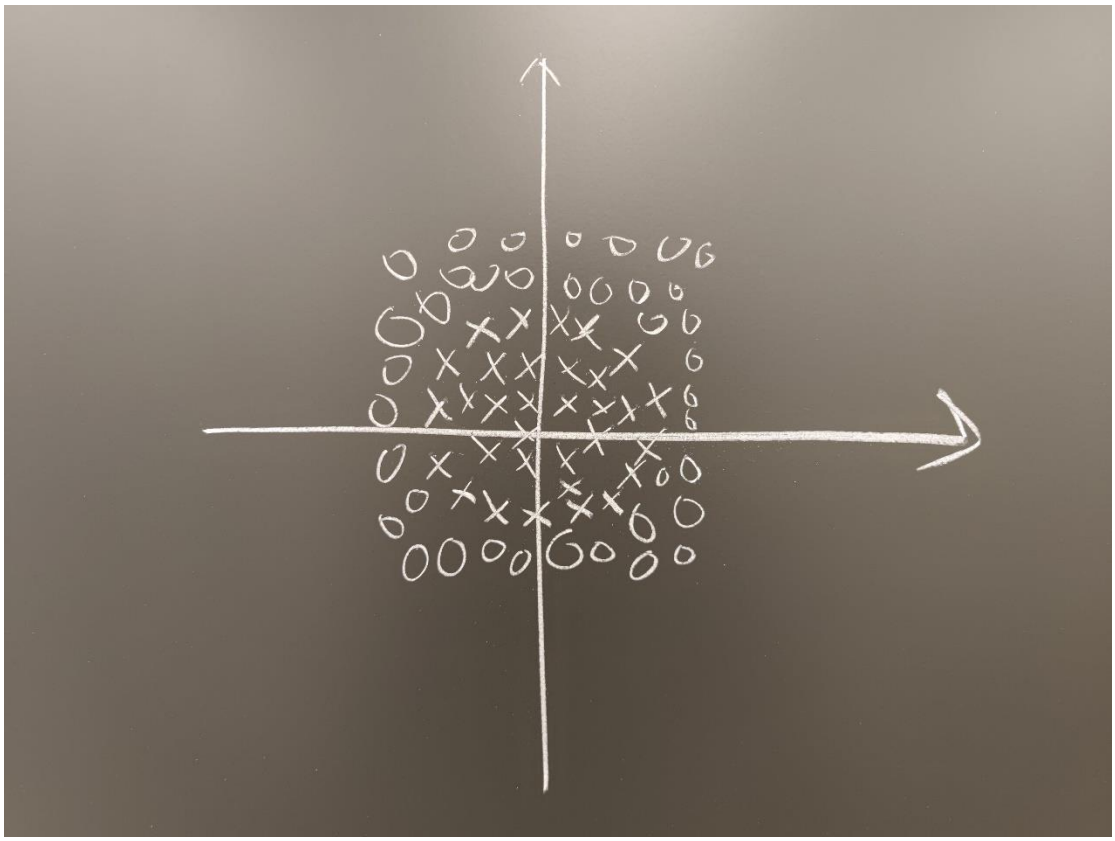
The margin is the distance between two separate hyperplanes (decision boundaries). This margin is determined by the support vector, which is the data point closest to the decision boundary. The optimization goal of SVM is to maximize this margin to improve the accuracy and robustness of classification.

(d) What is a naive Bayes classifier?

Naive Bayes classifier is a classification method based on Bayes' theorem and the independent assumption of features. By calculating the conditional probability of the items to be classified under each category, the category with the largest conditional probability is selected as the category of the items to be classified. Naive

Bayes classifiers assume that each feature of the sample is uncorrelated with the others. So it's called naive Bayes classification.

(e) Give an example of a classification problem in $\mathbb{R}^2$ that cannot be solved by a linear threshold function in the original input space.



A linear line cannot separate the cross mark from the circle mark, only the circle can.

Problem 2

Since

$P(y=0)=\frac{1}{2}$

$P(y=1)=\frac{1}{2}$

$P(x|y=0) \sim N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$P(x|y=1) \sim N(0,2) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{x^2}{8}}$

By Bayes' theorem:

$$P(y=0|x) = \frac{P(x|y=0)P(y=0)}{P(x)} = \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{2}}{P(x)}$$

$$P(y=1|x) = \frac{P(x|y=1)P(y=1)}{P(x)} = \frac{\frac{1}{2\sqrt{2\pi}} e^{-\frac{x^2}{8}} \cdot \frac{1}{2}}{P(x)}$$

The Bayes optimal classifier will assigns x to class 0 when $P(y=0|x) > P(y=1|x)$ otherwise assigns x to class 1.

$$P(y=0|x) > P(y=1|x) \Rightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} > \frac{1}{2\sqrt{2\pi}} e^{-\frac{x^2}{8}}$$

$$\Rightarrow e^{-\frac{x^2}{2}} > \frac{1}{2} e^{-\frac{x^2}{8}}$$

$$\Rightarrow x^2 < \ln(4)$$

So, the decision boundary is $x = \sqrt{\ln(4)}$

If $x < \sqrt{\ln(4)}$, assigns x to class 0.

If $x \geq \sqrt{\ln(4)}$, assigns x to class 1.

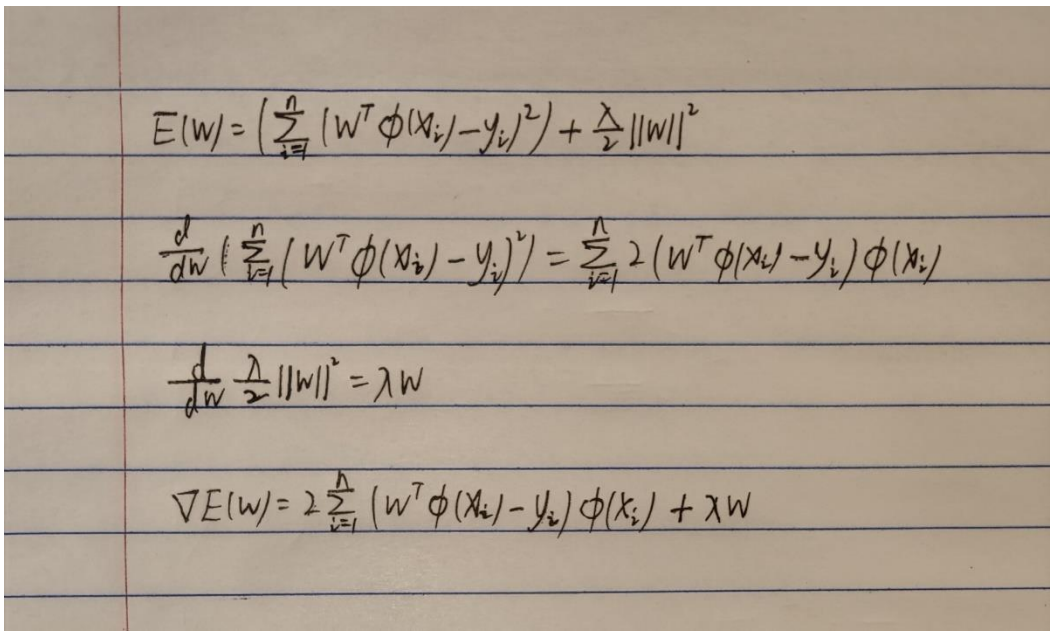
Problem 3

(a) Why is regularization useful and how does the value of λ affect the solution to the regression problem?

Because regularization prevents overfitting, it prevents the generation of complex models that are too close to the training data. Makes the model perfectly fit the training data but cannot generalize to previously unseen data. Regularization terms penalize larger values of model parameters. Regularization terms also encourage smaller model parameters, resulting in a smoother and more stable model. Because small parameter values are less sensitive to noise or outliers of the training data.

Larger λ values impose a stronger penalty on large parameter values, resulting in simpler models that are less prone to overfitting. Conversely, smaller λ values allow the model to be more complex, potentially improving the fit to the training data, but also increasing the risk of overfitting. λ controls the trade-off between how well the training data is fitted and ensuring the simplicity or smoothness of the model. By adjusting λ, we can find a model that achieves the best balance between these two goals.

(b) Give a formula for ∇E(w).



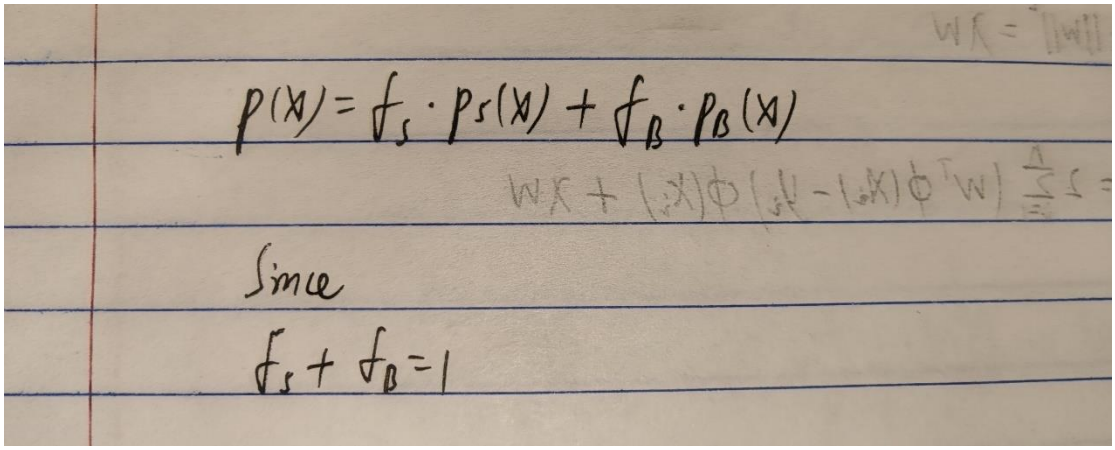
(c) Explain how gradient descent can be used to minimize E(w). You should give a clear description of the resulting algorithm with pseudo-code.

```
# Initialize the parameters
w = initialize_weights()
# Set a learning rate
learning_rate = 0.001
# Set the regularization parameter λ
```

```
regularization_parameter = 0.1
# Set the maximum number of iterations
iterations = 800


# Define the gradient function
# y is target label
def compute_gradient(X, y, w, lambda_value):
    return 2 * X.T.dot(X.dot(w) - y) + lambda_value * w


# Define the gradient descent function
def gradient_descent(X, y, w, learning_rate, lambda_value,
iterations):
    for iteration in range(iterations):
        gradient = compute_gradient(X, y, w, lambda_value)
        w = w - learning_rate * gradient
    return w

# Gradient descent
best_w = gradient_descent(X, y, w, learning_rate,
regularization_parameter, iterations)
```

Problem 4

(a) Consider the length x of random fish from the river. What is the density of x?



(b) Suppose we record the length of n random fish from the river. How can we use the recorded lengths $x_1, \ldots, x_n$ to estimate f_S and f_B without classifying the individual fish?