Aaron Vo

CPE 695

March 10, 2024

Homework 3

1. Explain what is the bias-variance trade-off? Describe few techniques to reduce bias and variance respectively.

The bias-variance tradeoff is when there is a tradeoff between the two types of error. A model with high bias can lead to underfitting where a complex data can be too simplified. A model with high variance can lead to overfitting, may fit the training data too closely, and fail to generalize to new data. The goal is to find the right balance that minimizes both bias and variance, resulting in a model that performs well on both training and new data.

To reduce bias, we can increase the model’s complexity or Include more relevant features or perform transformations on existing features to enhance the model's ability to represent the underlying patterns in the data.

In order to reduce variance, Introduce penalties for large coefficients in the model to prevent overfitting or use k-fold cross-validation to assess the model's performance on different subsets of the data.

1. Assume the following confusion matrix of a classifier. Please compute its precision, recall, and F1-score.

|  |  |  |
| --- | --- | --- |
|  | Class 1 | Class 2 |
| Class 1 | 50 | 30 |
| Class 2 | 40 | 60 |

Precision = 50 / (50 + 40) = .556

Recall = 50 / (50 + 30) = .625

F1-score = 2 \* (Precision \* Recall) / (Precision + Recall) = .5885

1. The naïve Bayes method is an ensemble method as we learned in Module 5. Assuming we have 3 classifiers, and their predicted results are given in the table 1. The confusion matrix of each classifier is given in table 2. Please give the final decision using the Naïve Bayes method:

|  |  |
| --- | --- |
| Sample X | Result |
| Classifier 1 | Class 1 |
| Classifier 2 | Class 1 |
| Classifier 3 | Class 2 |

|  |  |  |
| --- | --- | --- |
|  | Class 1 | Class 2 |
| Class 1 | 40 | 10 |
| Class 2 | 30 | 20 |

|  |  |  |
| --- | --- | --- |
|  | Class 1 | Class 2 |
| Class 1 | 20 | 30 |
| Class 2 | 20 | 30 |

|  |  |  |
| --- | --- | --- |
|  | Class 1 | Class 2 |
| Class 1 | 20 | 30 |
| Class 2 | 20 | 30 |

**Sample 1**

Classifier 1:

* P(Class 1 | Classifier 1) = 40 / (40 + 10) = 0.8
* P(Class 2 | Classifier 1) = 10 / (40 + 10) = 0.2

Classifier 2:

* P(Class 1 | Classifier 2) = 20 / (20 + 30) = 0.4
* P(Class 2 | Classifier 2) = 30 / (20 + 30) = 0.6

Classifier 3:

* P(Class 1 | Classifier 3) = 50 / (50 + 0) = 1.0
* P(Class 2 | Classifier 3) = 0 / (50 + 0) = 0.0

product of these probabilities for each class:

* Class 1: 0.8 \* 0.4 *\** 1.0 = 0.32
* Class 2: 0.2\* 0.6 \*0.0 = 0.0

Class 1 because it has a higher probability.

**Sample 2**

Classifier 1:

* P(Class 1 | Classifier 1) = 30 / (30 + 20) = 0.6
* P(Class 2 | Classifier 1) = 20 / (30 + 20) = 0.4

Classifier 2:

* P(Class 1 | Classifier 2) = 20 / (20 + 30) = 0.4
* P(Class 2 | Classifier 2) = 30 / (20 + 30) = 0.6

Classifier 3:

* P(Class 1 | Classifier 3) = 40 / (40 + 10) = 0.8
* P(Class 2 | Classifier 3) = 10 / (40 + 10) = 0.2

product of these probabilities for each class:

* Class 1: 0.6 \* 0.4 \* 0.8 = 0.192
* Class 2: 0.4 \* 0.6 \* 0.2 = 0.048

Class 1 because it has the higher probability.

**Sample 3**

For the third sample, the calculations are as follows:

Classifier 1:

* P(Class 1 | Classifier 1) = 0 / (0 + 50) = 0.0
* P(Class 2 | Classifier 1) = 50 / (0 + 50) = 1.0

Classifier 2:

* P(Class 1 | Classifier 2) = 0 / (0 + 50) = 0.0
* P(Class 2 | Classifier 2) = 50 / (0 + 50) = 1.0

Classifier 3:

* P(Class 1 | Classifier 3) = 1.0
* P(Class 2 | Classifier 3) = 0.0

Both probabilities will be zero so we will pick neither.