1 Maximum Likelihood Estimation

Recall that a Geometric distribution is defined as the number of Bernoulli trials needed to get one success. $P(X = k) = p(1 - p)^{k-1}$.

We observe the following samples from a Geometric distribution:

$x_1 = 5, x_2 = 8, x_3 = 3, x_4 = 5, x_5 = 7$

What is the maximum likelihood estimate for $p$?

2 Naive Bayes

In this question, we will train a Naive Bayes classifier to predict class labels $Y$ as a function of input features $A$ and $B$. $Y$, $A$, and $B$ are all binary variables, with domains 0 and 1. We are given 10 training points from which we will estimate our distribution.

<table>
<thead>
<tr>
<th>A</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Y</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(a) What are the maximum likelihood estimates for the tables $P(Y)$, $P(A|Y)$, and $P(B|Y)$?

| $Y$ | $P(Y)$ | $A$ | $Y$ | $P(A|Y)$ | $B$ | $Y$ | $P(B|Y)$ |
|-----|--------|-----|-----|---------|-----|-----|---------|
| 0   | 0      | 0   | 0   | 1       | 0   | 0   |         |
|     |        | 1   | 0   |         | 1   | 0   |         |
|     |        | 0   | 1   |         | 0   | 1   |         |
| 1   |        | 1   | 1   |         | 1   | 1   |         |

(b) Consider a new data point ($A = 1, B = 1$). What label would this classifier assign to this sample?

(c) Let’s use Laplace Smoothing to smooth out our distribution. Compute the new distribution for $P(A|Y)$ given Laplace Smoothing with $k = 2$.

| $A$ | $Y$ | $P(A|Y)$ |
|-----|-----|---------|
| 0   | 0   |         |
| 1   | 0   |         |
| 0   | 1   |         |
| 1   | 1   |         |

(a) What is the minimum number of parameters needed to fully model a joint distribution \( P(Y, F_1, F_2, ..., F_n) \) over label \( Y \) and \( n \) features \( F_i \)? Assume binary class where each feature can possibly take on \( k \) distinct values.

(b) Under the Naive Bayes assumption, what is the minimum number of parameters needed to model a joint distribution \( P(Y, F_1, F_2, ..., F_n) \) over label \( Y \) and \( n \) features \( F_i \)? Assume binary class where each feature can take on \( k \) distinct values.

(c) You suspect that you are overfitting with your Naive Bayes with Laplace Smoothing. How would you adjust the strength \( k \) in Laplace Smoothing?

- Increase \( k \)
- Decrease \( k \)

(d) While using Naive Bayes with Laplace Smoothing, increasing the strength \( k \) in Laplace Smoothing can:

- Increase training error
- Increase validation error
- Decrease training error
- Decrease validation error

(e) It is possible for the perceptron algorithm to never terminate on a dataset that is linearly separable in its feature space.

- True
- False

(f) If the perceptron algorithm terminates, then it is guaranteed to find a max-margin separating decision boundary.

- True
- False

(g) In binary perceptron where the initial weight vector is \( \vec{0} \), the final weight vector can be written as a linear combination of the training data feature vectors.

- True
- False

(h) For binary class classification, logistic regression produces a linear decision boundary.

- True
- False

(i) In the binary classification case, logistic regression is exactly equivalent to a single-layer neural network with a sigmoid activation and the cross-entropy loss function.

- True
- False

(j) You train a linear classifier on 1,000 training points and discover that the training accuracy is only 50%. Which of the following, if done in isolation, has a good chance of improving your training accuracy?

- Add novel features
- Train on more data

(k) You now try training a neural network but you find that the training accuracy is still very low. Which of the following, if done in isolation, has a good chance of improving your training accuracy?

- Add more hidden layers
- Add more units to the hidden layers