Name: Pratik Jasani

Course: 605.645

**K – Nearest Neighbors**

|  |  |  |  |
| --- | --- | --- | --- |
| # | x1 | X2 | Y |
| 1 | 0.23 | 0.81 | 0.18 |
| 2 | 0.42 | 0.78 | 0.33 |
| 3 | 0.64 | 0.23 | 0.14 |
| 4 | 0.87 | 0.19 | 0.17 |
| 5 | 0.76 | 0.43 | 0.32 |
| 6 | 0.39 | 0.63 | ? |

1. **What are the 3 nearest neighbors to Point 6? What is the predicted value of y for Point 6?**

* The calculations below

**(1-6) -** (0.39-0.23) ^2 + (0.63-0.81)^2 = 0.058

**(2-6) -** (0.39-0.42) ^2 + (0.63-0.78)^2 = 0.023

**(3-6) -** (0.39-0.64) ^2 + (0.63-0.23)^2 = 0.2225

**(4-6) -** (0.39-0.87) ^2 + (0.63-0.19)^2 = 0.424

**(5-6) -** (0.39-0.76) ^2 + (0.63-0.43)^2 = 0.1769

The three nearest neighbors are points 2, 5, 1. The predicted value of y for point 6 is 0.263.

Y for point 6 is calculated by (0.33+0.14+0.32)/3 = 0.2633

1. **Do you need to take the square root to compare distances?**

* We do not need to take the square root because for comparison it does not change anything.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Actual | |
|  |  | Positive | Negative |
| Predicted | Positive | 329 | 35 |
| Negative | 87 | 357 |

1. **What is the accuracy?**

* (329+357) / (329+357+87+35) =0.849

1. **What is the error?**
2. (87+35) / (329+357) = 0.17784
3. **What is the precision?**

* 329 / (329 + 87) = 0.7909

1. **What is the recall?**

* 329 / (329 + 35) = 0.90385

**Mean Squared Error**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **y** | **Y\_hat** | **(y-y\_hat)** | **(y-y\_hat)^2** | **(y-y\_mean)^2** |
| 3.78 | 3.32 | 0.46 | 0.2116 | 0.0605 |
| 4.82 | 5.21 | -0.39 | 0.1521 | 1.6538 |
| 2.83 | 2.97 | -0.14 | 0.0196 | 0.495616 |
| 2.76 | 2.37 | 0.39 | 0.1521 | 0.5999 |
| 3.48 | 3.29 | 0.19 | 0.0361 | 0.002916 |

1. **What is the MSE of the model?**

* (1/5) \* (0.2116+0.1521+0.0196+0.121+0.0361) = 0.10808

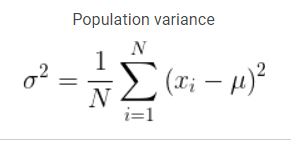
1. **What is the mean of y?**

* (1/5) \* (3.78+4.82+2.83+2.76+3.48) = 3.534

1. **If you used the mean of y as a predictor of y (y\_hat), what would its MSE be?**

* (1/5) \* (0.0605+1.6538+0.495616+0.5999+0.002916)= 0.5625464

1. **Look up for the formula for variance of a population. Surprised?**



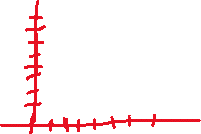
**Learning Curves**

**Give an example of a set of learning curves where collecting more data may help**

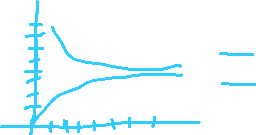
1. **Give an example of a set of learning curves where collecting more data may help.**

**Test**

**Train**



1. **Give an example of a set of learning curves where collecting more data probably will not help.**



**Test**

**Train**

