Machine Learning

Avik Bag

20th January 2017

1 Theory Questions

1. Question 1

$$
\text{Raw data} = \begin{bmatrix}
-2 & 1 \\
-5 & -4 \\
-3 & 1 \\
0 & 3 \\
-8 & 11 \\
-2 & 5 \\
1 & 0 \\
5 & -1 \\
-1 & -3 \\
6 & 1
\end{bmatrix}
$$

First we need to standardize this dataset. This is done by finding the mean and standard deviation along the column. After that, the mean is subtracted from the respective column and then divided by the corresponding standard deviation.

Average $(\mu) = \begin{bmatrix} -0.9 & 1.4 \end{bmatrix}$ $StandardDeviation(\sigma) = \begin{bmatrix} 4.228 & 4.273 \end{bmatrix}$

$$
\text{Standardized data} = \begin{bmatrix}
-0.2602 & -0.0936 \\
-0.9697 & -1.2635 \\
-0.4967 & -0.0936 \\
0.2129 & 0.3744 \\
-1.6792 & 2.2462 \\
-0.2602 & 0.8423 \\
0.4494 & -0.3276 \\
1.3954 & -0.5615 \\
-0.0237 & -1.0295 \\
1.6319 & -0.0936
\end{bmatrix}
$$

As per the lecture slides, since the data is now centered, the following equation can be used to calculate the covariance matrix based on the standardized data.

Covariance Matrix= $\frac{X^T X}{N-1}$

2. Question 2