**P1-Q1**

Basically, this problem does not have a correct answer. The motivation of this problem is to understand how each of these algorithms can be used to output a confidence value.

Naïve Bayes is preferable since we can still use it for classification when the values of some of the features are unknown.

In decision tree, since it gives us interpretable answer, it also can be preferable.

In logistic regression, it has a linear decision boundary so it is sensitive to add up features. However, it can give reliable results.

P1-Q2

P2-Q1

|  |  |  |  |
| --- | --- | --- | --- |
|  | ACC | Sens | Spe |
| GNB | 0.86 | 0.78 | 0.93 |
| Decision Tree | 0.72 | 0.63 | 0.79 |
| Logistic Regression | 0.79 | 0.75 | 0.82 |

In terms of the metrics, GNB is the best.

P2-Q2

|  |  |  |  |
| --- | --- | --- | --- |
|  | ACC | Sens | Spe |
| GNB | 0.98 | 1.0 | 0.98 |
| Logistic Regression | 1.0 | 1.0 | 1.0 |
| SVM | 0.98 | 1.0 | 0.98 |

**P2-Q3**



This is a bar chart of estimated coefficients of features (variables) over all features (variables).

Indices of non-zero coefficients are {12,26,65,70,118,126} under default parameter lambda (=1.0).

P3

|  |  |  |  |
| --- | --- | --- | --- |
|  | ACC | Sens | Spe |
| GNB | 0.74 | 0.71 | 0.78 |
| Decision Tree | 0.71 | 0.69 | 0.73 |
| Logistic Regression | 0.74 | 0.73 | 0.74 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | ACC | Sens | Spe |
| GNB | 0.96 | 0.98 | 0.95 |
| Logistic Regression | 0.96 | 0.94 | 0.97 |
| SVM | 0.95 | 0.98 | 0.93 |