Part 1

1.

$$k(X_i, X_j) = \sum_{d=1}^{D} X_{i,d} X_{j,d}$$

$$k(X_i, X_j) = \phi(X_i)\phi(X_j)^T$$

$$\phi(X_i)\phi(X_j)^T = \sum_{d=1}^{D} X_{i,d} X_{j,d} = X_i X_j^T$$

$$\phi(u) = u$$

For example,

$$X_i = [a_1 \quad a_2 \quad a_3]$$

$$X_j = [b_1 \quad b_2 \quad b_3]$$

$$k(X_i, X_j) = a_1 b_1 + a_2 b_2 + a_3 b_3$$

$$\phi(X_i)\phi(X_j)^T = X_i X_j^T = a_1 b_1 + a_2 b_2 + a_3 b_3$$

2.

False.

When number of features are greater than samples, the data is linearly separable, and we do not have to let it go to higher detentions. Thus, Gaussian Kernel is not better than linear kernel.

Part 2

Precision: 0.914821

Recall: 0.913265

F-Measure: 0.914043

Accuracy: 0.934116

Part 3

Accuracy: 0.906780