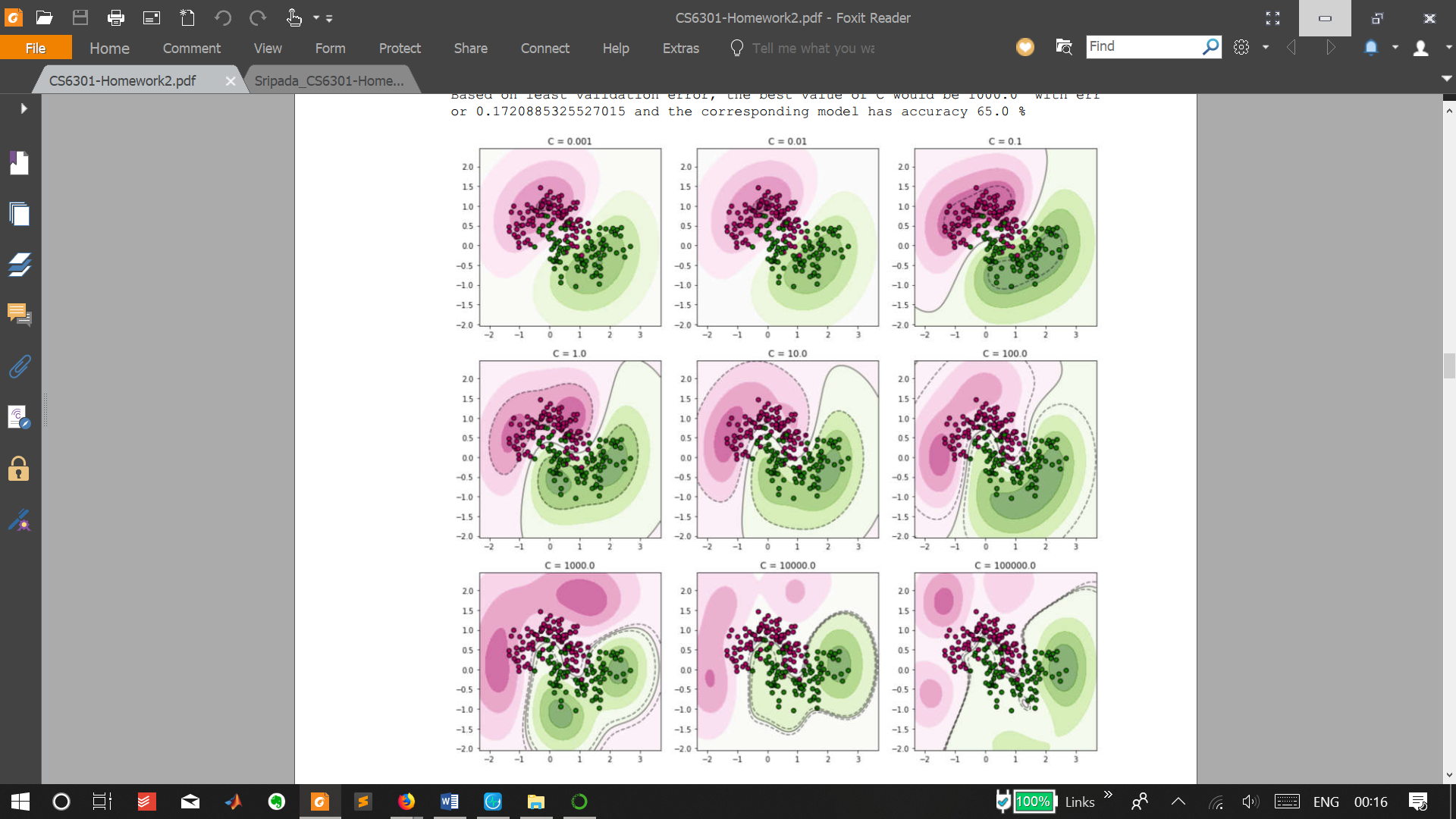
CS6301.010: Machine Learning for Engineers and Scientists – HW2 – RBF-SVMs

Siva Saket Sripada (2021432772)

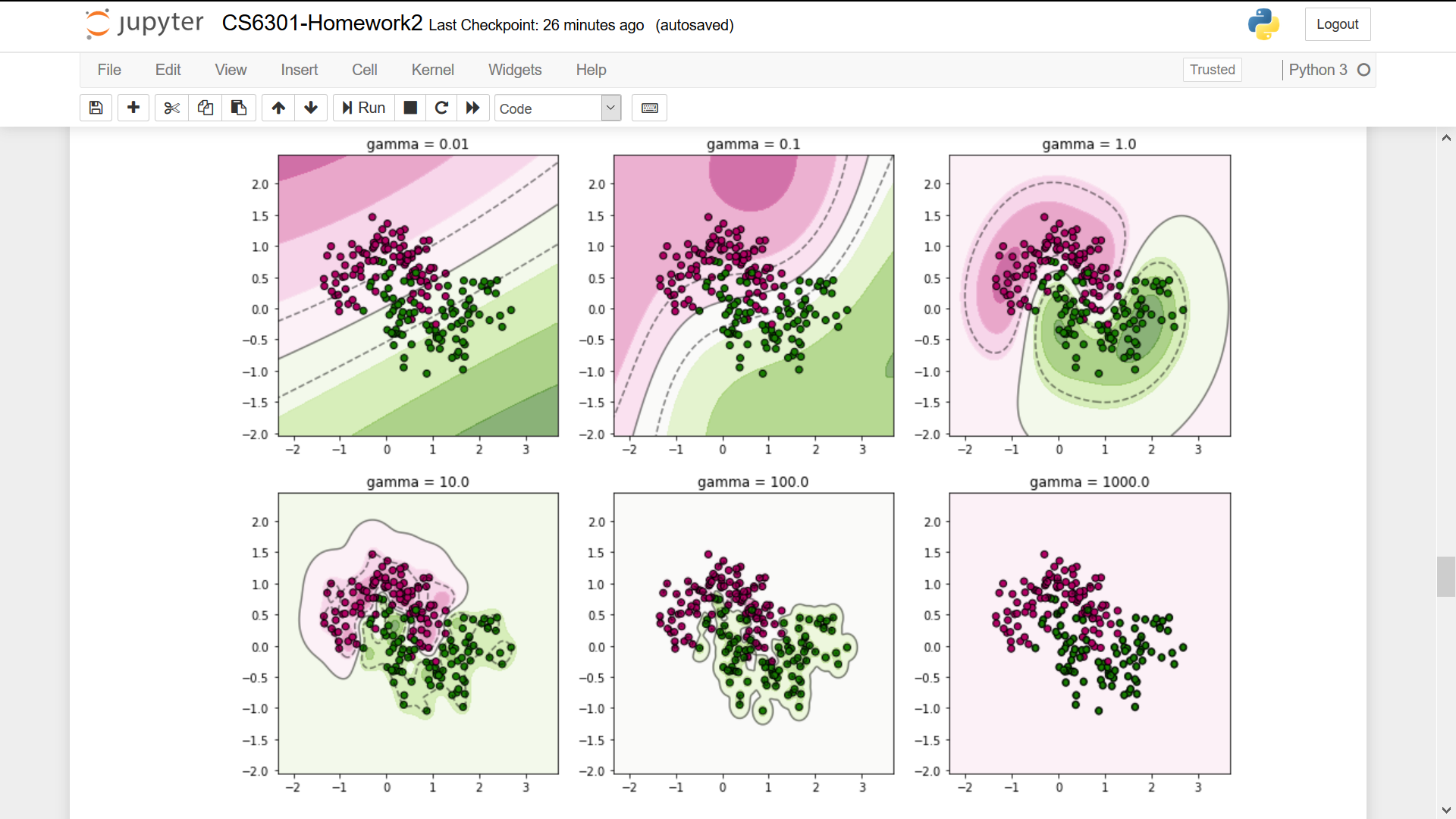
Objective: To understand the behaviour of SVMs with Radial-Basis Function (RBF) kernels with different values of C and gamma.

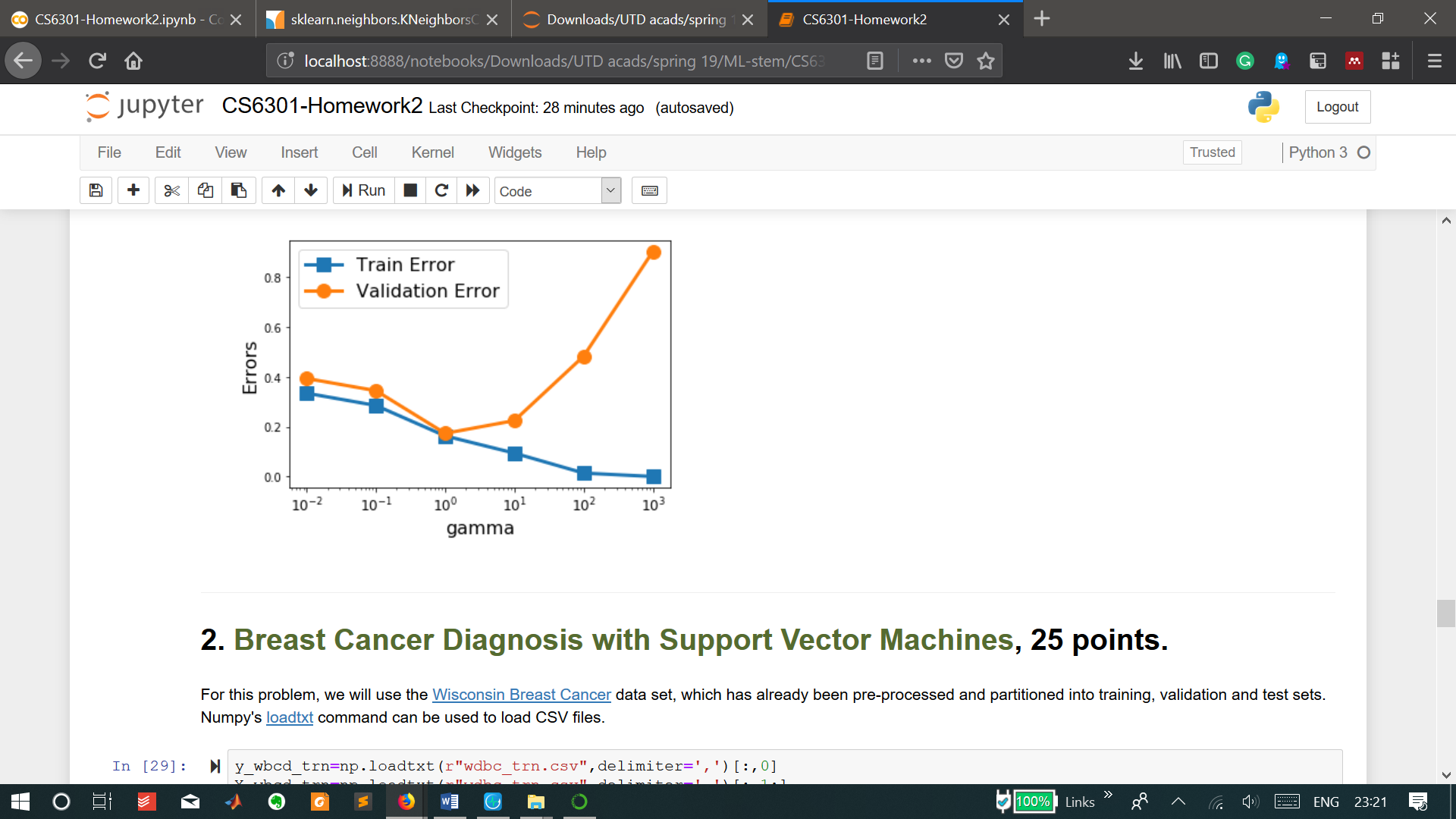
Q1

Varying C with gamma set as ‘scale’ :

1. Training error continuously decreses with increasing C – this is cooresponding to higher weightage to minmising the term corresponding to hinge-loss as compared to the term for maximising margin in the objective function of SVM. This is expected since the model will more aggressively minimise error and care lesser about margin.
2. Validation error follows a similar trend as training error but the validation error starts rising for higher values of C and blows up beyond 107.
3. Looking at the validation error curve and closely at the visualisations of classifiers indicate strongly that C in the range of 1-100 as optimal simply because:
   1. Low values of C yield classifiers that aggressively maximise the margin and separation but accommodate higher errors. This also leads to a more linear-hyperplane-separator-like behaviour as is seen in the cases with C=0.001, 0.01. This is essentially underfitting.
   2. High values of C yield models that aggresively miminise error and overfor the data while ignoring the separation between the classes, making it prone to erroneous future prediction – as can be seen in the classifier-visualisation as absurd boundaries.
   3. Among C=1, 10, and 100, errors are comparable and low, but visually the examples are best classified by C=10 as the boundaries for the two classes enompass all the examples and also the validation error is least (0.18). These values yield models that strike a balance between margin maximisation (fit) and error minimisation (complexity).
4. Based on least validation error though,the algorithm picked C=1000 as the best value based on the validation error of 0.17, but that classifaction looks off.
5. Hence I pick C=10 as the best choice for penalty-parameter,C.

Q2

Varying gamma with fixed C=10:

1. Training error decreseas with increasing gamma because of how the kernel transform works… it is a negative exponent which decays to zero rapidly. So a high gamma value would diminish the effect of similarity between two examples making the predictions erroneous. This is evident since high gamma values yield models with blowing-up validation error. (High gammas implies aggressively minimising complexity)
2. Low gamma means larger influence of similarity between two vectors on the classification – higher complexity and bias and errors too, as can be seen with higher validation error.
3. gamma=1 strikes a healthy balance between over and undermining the similarity of two vectors and hence between fit and complexity in a sense, thereby resulting in minimising error, which can be seen in the visualisation of classifiers

Q2

Similar to Q1, validation and test errors were obtained for different RBF-kernel SVM models with varying C and gamma on the Wisconsin Breast cancer data. The pair with the least validation error was picked as the best pair (as in Q1) and the resulting accuracy on test data was 96.52%.

Q3

kNN based on kdTree algorithm was used to model the same data as in Q2 for different values of k, and as in Q1 least validation error was used to pick k=1 as the best parameter which yielded 96.52% accuracy on test data.

I would pick an RBF-kernel-SVM over a kNN for this classification problem as the separating hyperplane would be smoother, the class boundaries well separated from the hyperplane and also since SVM handles outliers better than kNN (latter highly sensitive to outliers ad hence not preffered).

Additionally kNN’s classes would be jagged and hence more prone to erroneous predictions on future data even though accuracy on existing data is same as SVM.