**Assignment 4**

Chart, scatter chart

Description automatically generated

**1. Support Vector Machines with Synthetic Data**

**a. The effect of the regularization parameter C**

**Plot:**

Chart, line chart

Description automatically generated

**Discussion:**

Training Error: Training Error monotonically decreases while C increases. It decreases quickly at first and slowly in the end.

Validation Error: Validation Error decreases when C <= 1 and generally increases afterwards (overfitting).

A picture containing calendar

Description automatically generated

C is a trade-off between training error and flatness. Some people called C as Cross-validation parameter. While C increases, the model is softer, which means more slack are allowed and whole model is more precise and less misclassifying. Vice versa.

**Final Model Selection:**

Chart, line chart

Description automatically generated

To have the least Validation Error, Cbest = 1.

Console:

A picture containing text

Description automatically generated

**b. The effect of RBF kernel parameter 𝛄**

**Plot:**

Chart, line chart

Description automatically generated

**Discussion:**

Training Error: Training Error monotonically decreases while 𝛄 increases. It decreases quickly at first and slowly in the end.

Validation Error: Validation Error decreases when 𝛄 <= 1 and generally increases afterwards (overfitting).

Chart, calendar

Description automatically generated

𝛄 defines how far the influence of a single training example reaches. For a big 𝛄, it will generate a sharp heap which will locate most of its contribution near the center. Hence, less constrain will cause the model loss the sense of the overall shape of data. When 𝛄 is large enough, the model’s accuracy is close to 1 but useless for classification.

**Final Model Selection:**

Chart, line chart

Description automatically generated

To have the least Validation Error, 𝛄best = 1.

Text

Description automatically generated

**2. Breast Cancer Diagnosis with Support Vector Machines**

**Print Errors:**

Table

Description automatically generated

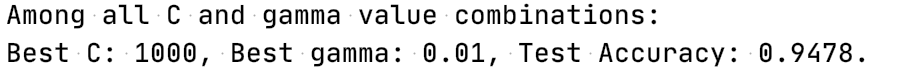
I use median to blur the matrix and find the final “best” (C, 𝛄) pair.

Background pattern

Description automatically generated

**Final Model Selection:**

To have the least Validation Error, Cbest = 1000 and 𝛄best = 0.01.



**3. Breast Cancer Diagnosis with k-Nearest Neighbors**

**Plot:**

Chart, line chart

Description automatically generated

**Final Model Selection:**

To have the least Validation Error, kbest = 5.

Graphical user interface, text

Description automatically generated

**Discussion:**

Depending on the result which I got, I will prefer to use kNN. For the test accuracy of kNN in kbest is larger than that of SVM in (Cbest, 𝛄best). But all these two is good for Breast Cancer Diagnosis:

Test Accuracy(SVMs(Cbest=1000, 𝛄best=0.01)) = 0.9478 < 0.9565 = Test Accuracy(kNN(kbest=5))