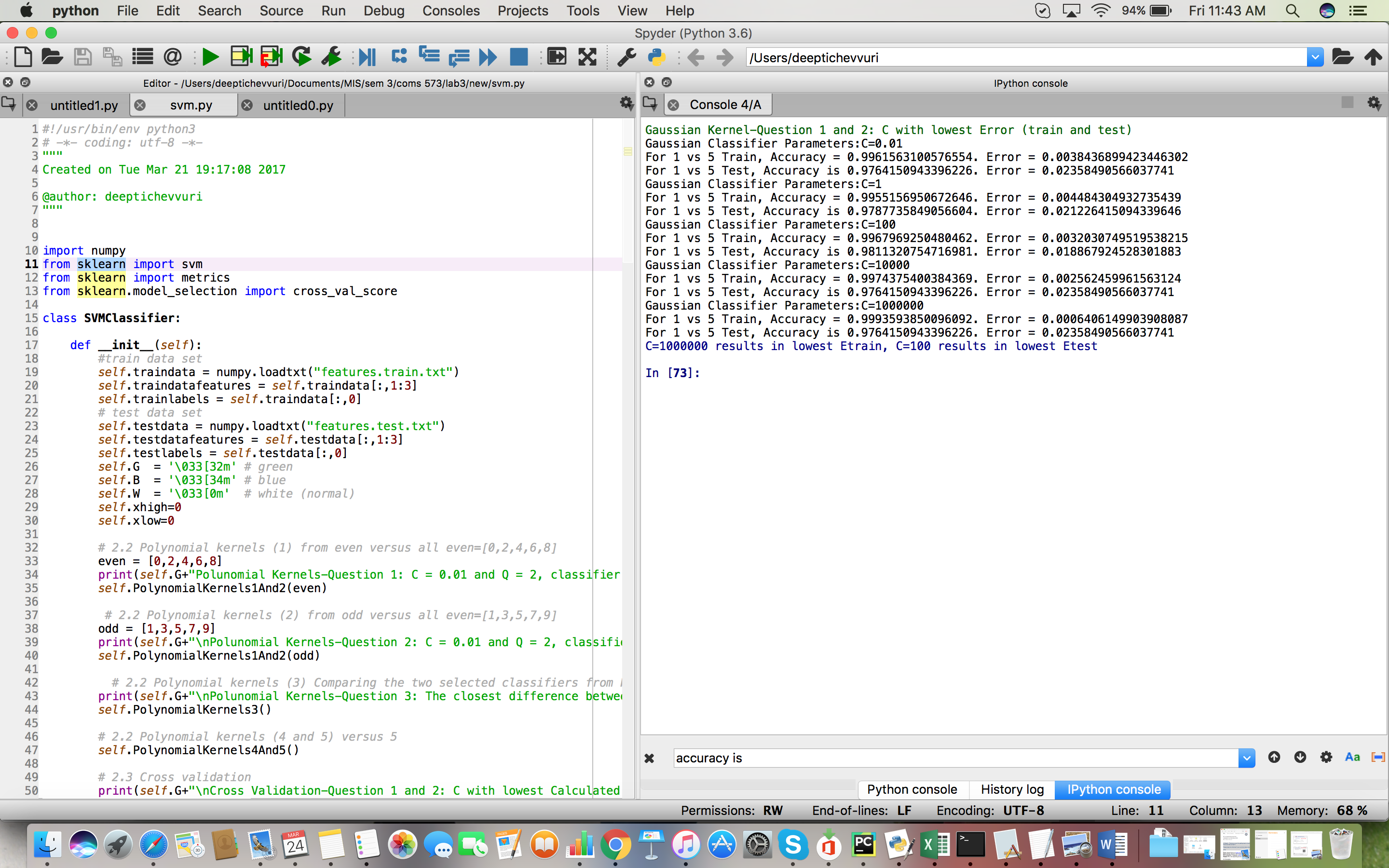
**Lab 3 Report:**

Please use Python3, I have used “sklearn” package with python 3 interpreters with Spyder IDE for this lab. (below screenshot)

Please keep the file in the same folder as the train and test text files while executing.

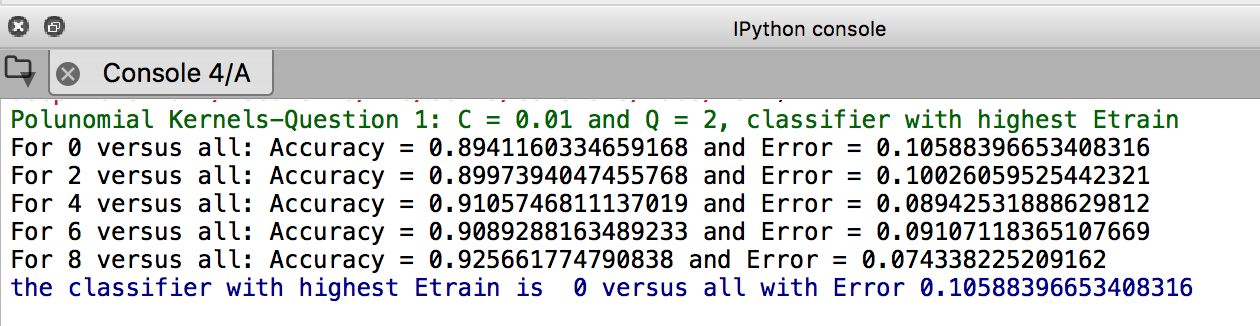
****

**2.1 Justification**

A Hard Margin SVM is a quadratic programming problem with d+1 variables, as the dimensionality of the Hard Margin Problem depends on the dimensionality of the feature space plus one variable to handle the over fitting problem.

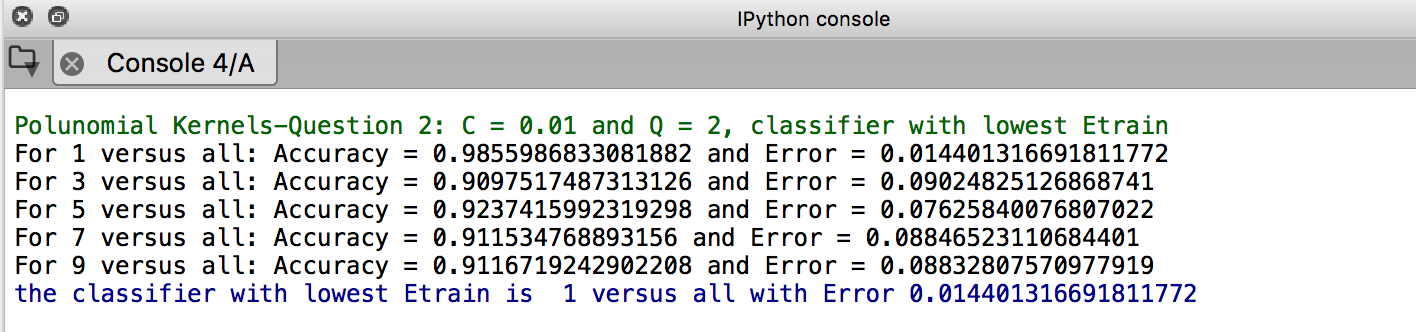
**2.2 Polynomial Kernels**

**1.** With C = 0.01 and Q = 2, which of the following classifiers has the highest Etrain



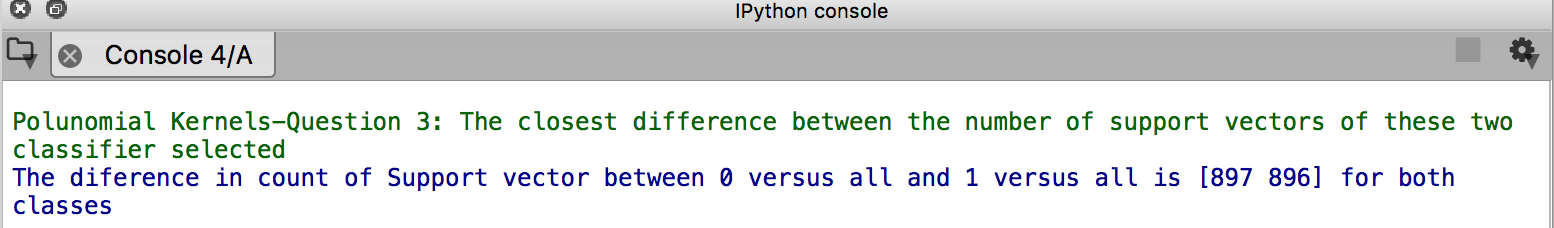
therefore, the answer is (a)

**2.** With C = 0.01 and Q = 2, which of the following classifiers has the lowest Etrain?



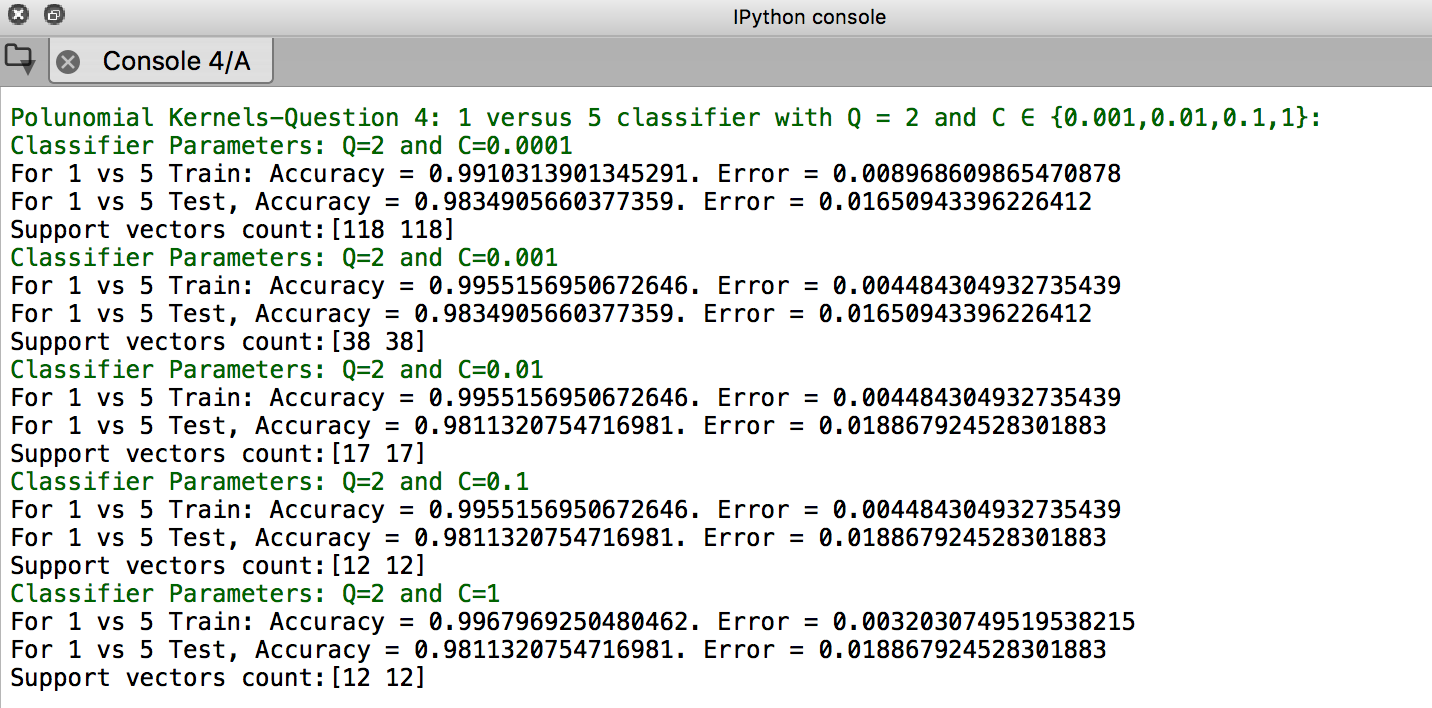
therefore, the answer is (b)

3. Comparing the two selected classifiers from Problems 1 and 2, which of the following values is the closest to the difference between the number of support vectors of these two classifiers?



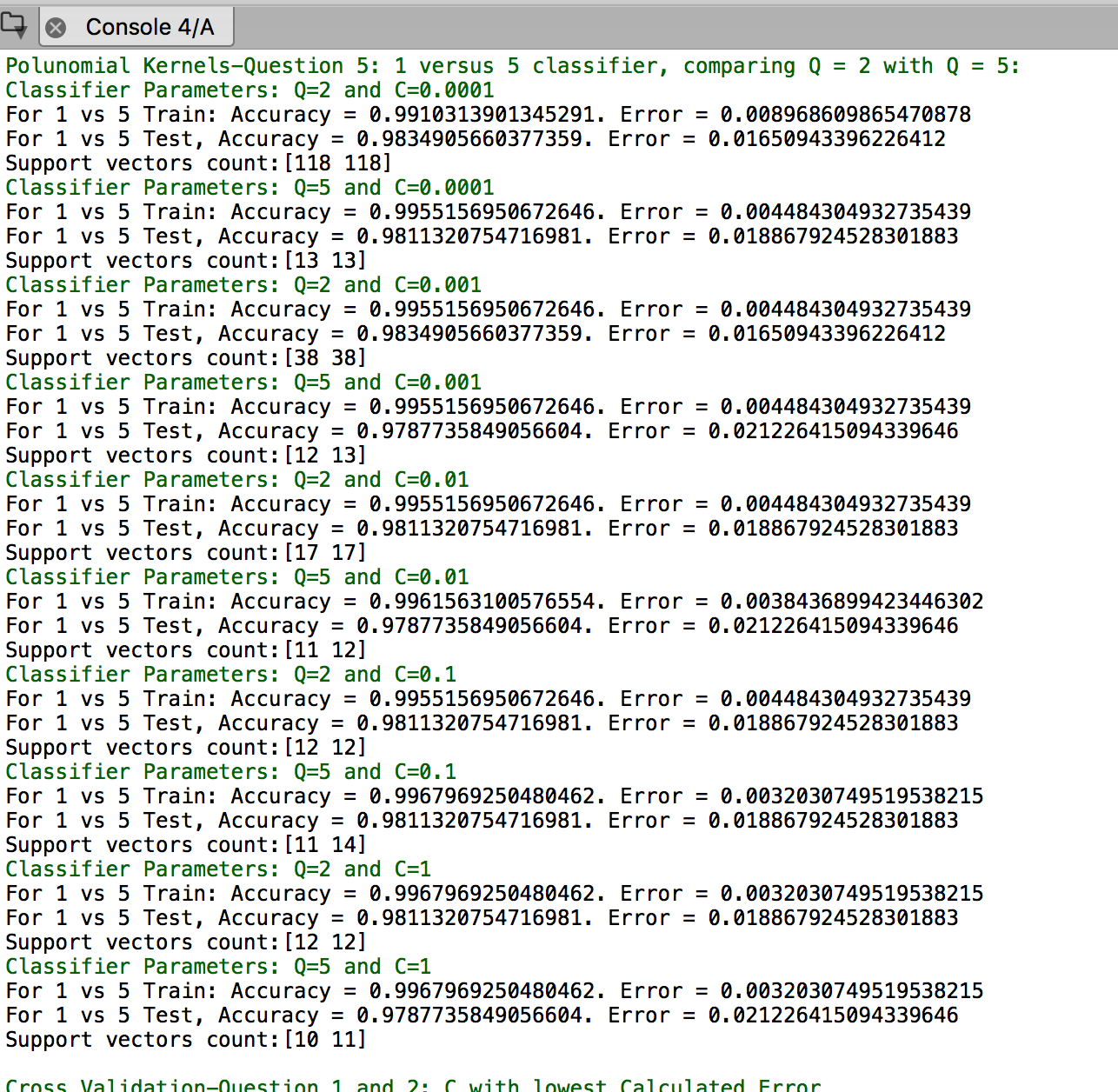
The sum of the difference between the support vectors for each class for the two classifiers is 897+896=1793, which means (c) 1800 is the correct answer.

4. Consider the 1 versus 5 classifiers with Q = 2 and C 2 {0.001,0.01,0.1,1}. Which of the following statements is correct? Going up or down means strictly so.



1. The number of support vectors goes down when C goes up. False (is 12 for both c=0.1 and c=1)
2. The number of support vectors goes up when C goes up. False
3. Etest goes down when C goes up. False
4. Maximum C achieves the lowest Etrain. True
5. None of the above. False

5. In the 1 versus 5 classifiers, comparing Q = 2 with Q = 5, which of the following statements is correct



(a) When C = 0.0001, Etrain is higher at Q = 5. False

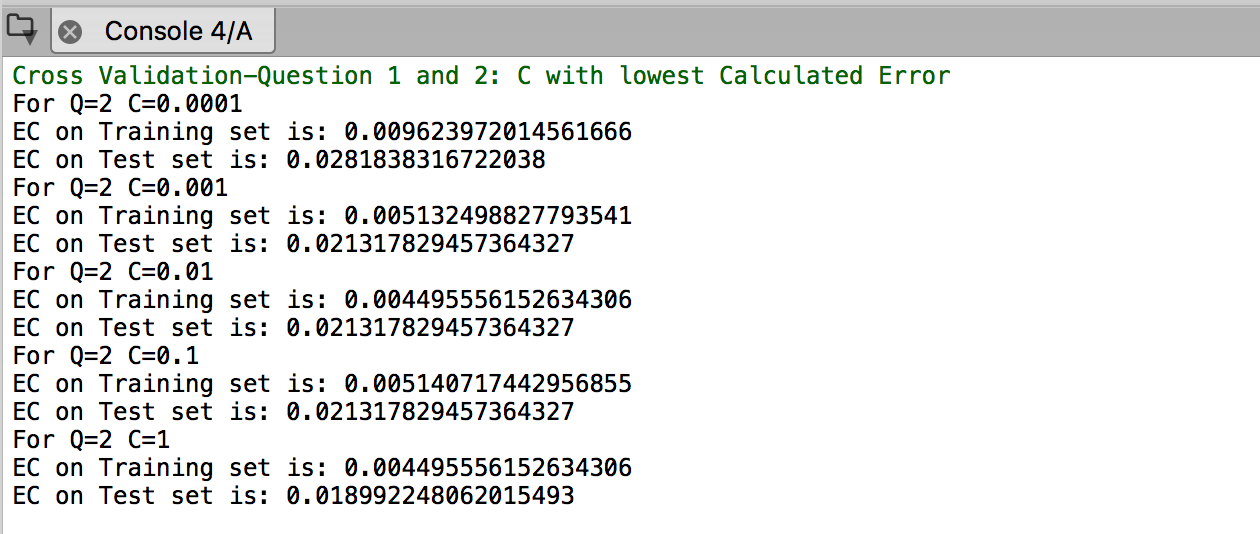
(b) When C = 0.001, the number of support vectors is lower at Q = 5. True

(c) When C = 0.01, Etrain is higher at Q = 5. False

(d) When C=1, Etest is lower at Q=5. False

(e) None of the above. False

**2.3 Cross Validation**



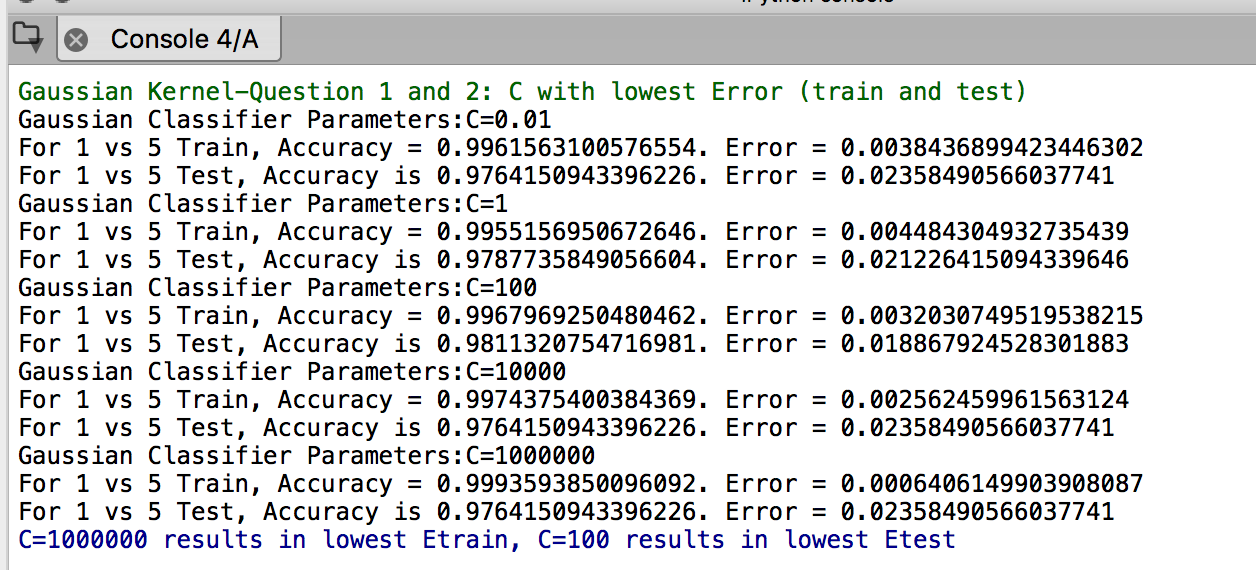
1. Consider the 1 versus 5 classifiers with Q = 2. We use Ecv to select C 2 {0.0001, 0.001, 0.01, 0.1, 1}. For each value of C, we run 10-fold cross validation on the training data and compute the Ecv over 100 runs. We should determine C with lowest Ecv. If there is a tie, select the smaller C. Which value of C shall you select?

Lowest Ecv (train) is for C=0.01 (ties between C=0.01 and C=1)

1. Again, consider the 1 versus 5 classifiers with Q = 2. For the winning selection in the previous  problem, the classification error on test set Etest is closest to

For the wining Lcaasifier in 1, Etest= 0.021317829457364327

* 1. **Gaussian Kernel**

****

1. Which of the following values of C results in the lowest Etrain? C=1000000

2. Which of the following values of C results in the lowest Etest? C=100