**Midterm Exam for Machine Learning, 4/23/2021**

Please show your work. That is, you need to provide the intermediate steps toward the answers. I cannot accept a reason like “the answer is directly obtained from a software package, so I don’t know how to calculate it myself.” It is OK to write a simple program with online resources such as Octave (an open source version of MATLAB) to save calculation time. But, you still need to hand-calculate at least one value to show how to use the equations.

1. The following table is a small portion of a big medical database to predict whether patients will need respiratory assistance.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Attributes | | | | | Class |
| Compliance | Resistance | Pulse oximeter | PCO2 | HCO3 |
| 6 | 10 | 97 | 23 | 22 | 2 |
| 100 | 11 | 100 | 29 | 22 | 2 |
| 64 | 13 | 100 | 37 | 24 | 1 |
| 51 | 12 | 100 | 25 | 18 | 1 |
| 33 | 16 | 100 | 40 | 18 | 1 |
| 34 | 12 | 100 | 37 | 25 | 2 |

If a particular patient has the following attributes, predict the class the patient will belong to by using a 3-NN classifier. For simplicity, no normalization is carried out. (15 points)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Compliance | Resistance | Pulse oximeter | PCO2 | HCO3 |
| 50 | 14 | 96 | 39 | 20 |

1. Follow problem 1, but this time uses a Naïve Bayesian classifier. Note: The attributes are not categorical, so you need to use Gaussian distributions in this problem. A simple program may help to reduce the computational burden. (15 points)
2. If one attribute in the training dataset in problem 1 is to be removed, which one is most likely to be removed? Why? (10 points)
3. If we want to use a decision tree to perform the prediction for the dataset in Problem 1, we need to discretize the continuous attributes. Follow the example in the lecture notes (16\_Decision trees) to find the optimal thresholds in C4.5 to binarize attributes of resistance and PCO2 into high and low. (10 points)
4. If we consider only the attribute of compliance in problem 1 and want to use the GMM to model this attribute with two mixtures. Suppose that the initial conditions are:

Update the parameters by completing the E and M steps in one epoch. (15 points)

1. Use the gradient descent method with penalty to compute and for the following: Minimize subject to . You may set the penalty constant to 1,000, the step size , and the initial condition . Based on your **,** and **,** do you think your answers move toward the right direction? Why? (15 points)
2. Please identify the support vectors of the following training data samples: and . Hint: Recall the meaning of support vectors. (10 points)
3. We use a special method to classify the iris dataset (see also Fig. below). The used discriminative classifiers are binary classifiers (meaning that it can only indicate whether the input sample is from class **+** or class **-)**. To extend the model to three classes, we use two identical models trained with different training sets. Then, during the test (prediction) phase, the final decision is made with the following rule:

**If** (model A predicts **+**) **then** class is Versicolor

**else** **if** (model B predicts **+**) **then** class is Setosa

**else** class is Virginica

Which classes of iris flowers should be included in the training set for the class “-” on model A? How about class “+” of model B? (10%)

Model A

Model B

Decision Rule

Test

Sample