

Homework 3, Due February 5

CAPP 30255 Advanced Machine Learning for Public Policy
University of Chicago

1. (30 points) In the following classification problem x_1 and x_2 are real-valued attributes:

x_1	x_2	Class Label
0.5	0.5	+
-0.5	0.5	+
-0.5	-0.5	+
2	2	-
2	-2	-
-2	2	-
-2	-2	-

Please solve the following questions by hand, showing your calculations clearly.

- (a) Roughly plot the position of the examples along x_1 and x_2 and choose a simple decision boundary such that all the seven training examples are correctly classified.
 - (b) Compute by hand the parameters of a logistic regression model that corresponds to the above decision boundary. (Hint: For features use x_1^2 , x_2^2 .)
 - (c) Determine the probability of the + class for the first and the last training examples according to the your logistic regression model.
2. (70 points) Write a program to implement a Naive Bayes classifier. Assume that all features are discrete valued, and there are only two classes.

For each of the attached datasets, compare the performance of your Naive Bayes classifier with the performance of the logistic classifier at [sklearn.linear_model.LogisticRegression](#). To compare performances, do the following:

- (a) Use [matplotlib.pyplot.contour](#) to plot the decision boundaries for the two classifiers.

- (b) Compare the F_1 scores for the two classifiers, obtained through 5-fold crossvalidation.
- (c) Discuss which of the two classifiers you would choose for the particular dataset.