

T-61.3050 Classification Challenge

Irene Kolomvrezou, Eduardo Castellanos

October 17, 2014

1 Brief description

1.1 Type of wine

In order to predict the wine type, we use the gradient descent method for multivariable classification. We perform cross-validation, splitting the data set into training set (70%) and validation set (30%). We are trying to find the right weights that will minimize our error function, so in every iteration we recalculate the derivative function of the error to derive smaller thetas until our error function converges to its minimum. The probability of a sample of being white or red is estimated using the sigmoid function and for our prediction we define that if the probability is larger than/equal to 0.5, then the sample is white, otherwise it's red.

1.2 Quality of wine

Trained a Gaussian MLE for each level, and each parameter. Thus resulting in generating 7x11x2 distributions for the classification. Each distribution was tested against the training set, and the ones from parameters that yielded the lowest error were selected for each quality level. This was then tested against the validation set so as to perform cross validation.