MTAT.03.227 Machine Learning
Practice session 1
Basics of linear classification

September 9-11, 2019

Exercise 1. Main concepts in classification.

Consider the task of deciding whether a pet is a cat or a dog, given its weight and height. To learn a classifier, we are given data about 5 cats and 5 dogs. The weights of the 5 cats are $4,5,5,5,6$ units and the heights are $2,1,2,3,2$ units. The weights of the 5 dogs are $7,11,11,13,13$ units and the heights are $4,8,10,8,10$ units. For convenience, let us call the weight units as *pettygrams* and the height units as *pettymeters*.

(a) How many instances are there? How many training instances? How many test instances? How many features? How many classes?

(b) What is the input space $\mathscr{X}$? What is the output space $\mathscr{Y}$?

(c) What is the first instance $\mathbf{x}_1$ here? Write it down in mathematical vector notation, that is $(\dots)$. What is its label y_1?

(d) Choose one of the classes to be called as positives and the other as negatives. Write down the contents of $\mathsf{Tr}^{\oplus}$ in mathematical set and vector notation, that is: $\{(\dots),(\dots),\dots\}$.

(e) Present the data as a single table.

(f) Visualise the data. Are the classes linearly separable? Can we use only one feature (weight or height) to separate the data?

Exercise 2. Basic linear classifier.

(a) Calculate and visualise the class centres $\mathbf{p}$ and $\mathbf{n}$.

(b) For any instance $\mathbf{x}$, the basic linear classifier predicts positive or negative, depending on whether $\mathbf{x}$ is closer to $\mathbf{p}$ or to $\mathbf{n}$. Does it predict the correct class on all training instances?

(c) Calculating distances requires taking squares and square roots, in what sense is this then a linear classifier?

(d) Find vector $\mathbf{w}$ and real number t, such that the basic linear classifier predicts positive if $\mathbf{w} \cdot \mathbf{x} > t$ and negative if $\mathbf{w} \cdot \mathbf{x} < t$. Apply this to demonstrate that one training instance is misclassified.

(e) What should the basic linear classifier predict for instances that are exactly on the decision boundary, that is $\mathbf{w} \cdot \mathbf{x} = t$? Draw the decision boundary.

(f) Does there exist a linear classifier which predicts correctly on all the training instances? Try to find one such classifier by changing t. If you can find one, then draw the new decision boundary.

Exercise 3. Perceptron.

(a) Represent the basic linear classifier learned on the cats-and-dogs data in homogeneous coordinates.

(b) Apply the perceptron algorithm on the training data, initialising with zero-vector and using learning rate $\eta = 1$. How many steps does it take to converge? Does the obtained model classify all training data correctly?

(c) What happens if we change the learning rate? What happens if we change the initial weights?