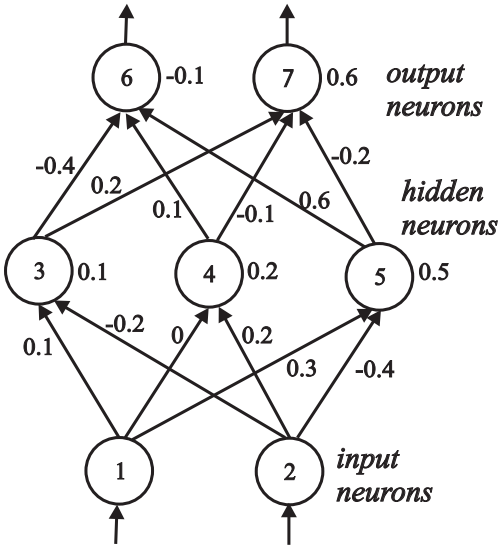
Tutorial exercises:
Backpropagation neural networks, Naïve Bayes, Decision Trees, k-NN, Associative Classification.

Exercise 1.

Suppose we want to classify potential bank customers as good creditors or bad creditors for loan applications. We have a training dataset describing past customers using the following attributes: Marital status {married, single, divorced}, Gender {male, female}, Age {[18..30[, [30..50[, [50..65[, [65+]}, Income {[10K..25K[, [25K..50K[, [50K..65K[, [65K..100K[, [100K+]}.
Design a neural network that could be trained to predict the credit rating of an applicant.

Exercise 2.

Given the following neural network with initialized weights as in the picture, explain the network architecture knowing that we are trying to distinguish between nails and screws and an example of training tupples is as follows: T1{0.6, 0.1, nail}, T2{0.2, 0.3, screw}.



Let the learning rate ŋ be 0.1 and the weights be as indicated in the figure above. Do the forward propagation of the signals in the network using T1 as input, then perform the back propagation of the error. Show the changes of the weights.

Exercise 3.

Why is the Naïve Bayesian classification called "naïve"?

Exercise 4. Naïve Bayes for data with nominal attributes

Given the training data in the table below (*Buy Computer* data), predict the class of the following new example using Naïve Bayes classification: age<=30, income=medium, student=yes, credit-rating=fair

RID	age	income	student	credit_rating	Class: buys_computer
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 . . . 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 . . . 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 . . . 40	medium	no	excellent	yes
13	31 . . . 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Exercise 5. Applying Naïve Bayes to data with numerical attributes and using the Laplace correction (to be done at your own time, not in class)

Given the training data in the table below (*Tennis* data with some numerical attributes), predict the class of the following new example using Naïve Bayes classification:
outlook=overcast, temperature=60, humidity=62, windy=false.

Tip. You can use Excel or Matlab for the calculations of logarithm, mean and standard deviation. The following Matlab functions can be used: `log2` – logarithm with base 2, `mean` – mean value, `std` – standard deviation. Type `help` <function name> (e.g. `help mean`) for help on how to use the functions and examples.

outlook	temperature	humidity	windy	play
sunny	85	85	false	no
sunny	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

Exercise 6. Using Weka (to be done at your own time, not in class)

Load iris data (iris.arff). Choose 10-fold cross validation. Run the Naïve Bayes and Multi-layer percepton (trained with the backpropagation algorithm) classifiers and compare their performance. Which classifier produced the most accurate classification? Which one learns faster?

***Exercise 7.* k-Nearest neighbours**

Given the training data in Exercise 4 (*Buy Computer* data), predict the class of the following new example using k-Nearest Neighbour for k=5: age<=30, income=medium, student=yes, credit-rating=fair. For distance measure between neighbours use a simple match of attribute values: distance(A,B)= $\sum_{i=1}^{4} w_i * \partial(a_i, b_i) \Big/ 4$ where $\partial(a_i, b_i)$ is 1 if a_i equals b_i and 0 otherwise. a_i and b_i are either *age, income, student* or *credit_rating*. Weights are all 1 except for income it is 2.

***Exercise 8.* Decision trees**

Given the training data in Exercise 4 (*Buy Computer* data), build a decision tree and predict the class of the following new example: age<=30, income=medium, student=yes, credit-rating=fair.

***Exercise 9.* Associative Classifier**

Given the training data in Exercise 4 (*Buy Computer* data), build an associative classifier model by generating all relevant association rules with support and confidence thresholds 10% and 60% respectively. Classify using this model the new example: age<=30, income=medium, student=yes, credit-rating=fair, selecting the rule with the highest confidence. What would be the classification if we chose to vote the class among all rules that apply?