Neural Networks KU WS19

Exercise Sheet 2

Linear classification [10 P + 2* P]

We consider the data set `vehicle.pkl`. The task is to classify a given silhouette as one of two types of vehicles, i.e., SAAB or VAN, using a set of features extracted from the silhouette. You are required to use Python for this assignment. The goal will be to minimize misclassification rate on the test set. The data has been splitted into a training set and a test set. First, extract from both of these sets only data points of the classes SAAB ($C = 2$) and VAN ($C = 4$).

a) [5 P] Use the *Probabilistic Generative Model* approach to classify the data set. Fit a probabilitstic generative model to the data. For the model, we assume Gaussian class-conditional distributions with a common covariance matrix. Estimate the class priors, the covariance matrix, and the means from the training data. The maximum likelihood estimates of the model parameters are given in the Appendix. Then classify using the posterior distribution over classes for both, the training set and the test set. Do the classification using only the first 2, then 3, etc. up to all features of the dataset. Plot the the classification accuracy (percentage of correctly classified examples) as a function of the number of input features for both, the training set and the test set. Also report any notable observations.

b) [5 P] Implement the *iterative reweighted least squares (IRLS)* algorithm and apply it to the dataset. Use a suitable stopping criterion. Plot the classification accuracy as in point (a) in dependence of the number of input features. Further, the report should contain plots showing the evolution of the cross-entropy error and the misclassification rate (training and test) over training epochs (iterations of IRLS) for a few of the above used feature-numbers.

   Report also the update rule used and the final performance in terms of the misclassification rate for all features. Compare the results to those of point a). The initial weights for IRLS have to be chosen with care. Choose them as small random numbers.

c) [2* P] For the case of two input features, plot the data-points in 2D and the decision boundaries for the generative model and IRLS.

Present your results clearly, structured, and legible. Submit your Python code and report at the TeachCenter. Do not zip the report together with the code. You can form groups of two students. Each student should submit a report (can be the same), with the team member indicated on the top right of the report.

Appendix

Let $\mathbf{x}^{(n)} \in \mathbb{R}^D$ denote the feature vector for the n-th training example. We assume a model with prior probability p_1 for class 1, $(1 - p_1)$ for class 2, and Gaussian class conditionals with means $\mu_1 \in \mathbb{R}^D$ and $\mu_2 \in \mathbb{R}^D$ and a shared covariance matrix Σ.

The Maximum Likelihood estimates of these parameters are

$$
\begin{aligned}
p_1 &= \frac{N_1}{N} \\
\mu_i &= \frac{1}{N_i} \sum_{n \in \mathcal{C}_i} \mathbf{x}^{(n)} \\
\Sigma &= \frac{N_1}{N} S_1 + \frac{N_2}{N} S_2 \text{ with} \\
S_i &= \frac{1}{N_i} \sum_{n \in \mathcal{C}_i} \left(\mathbf{x}^{(n)} - \mu_i\right) \left(\mathbf{x}^{(n)} - \mu_i\right)^T .
\end{aligned}
$$

Here, N denotes the number of training examples, N_i denotes the number of training examples for class C_i, and $\mathcal{C}_i$ denotes the set of indices for examples of class i.

The posterior distribution of this model was discussed in the lecture (see slides on decision theory).