Machine Learning
Exercise 12

Marc Toussaint
TAs: Janik Hager, Philipp Kratzer
Machine Learning & Robotics lab, U Stuttgart
Universitätsstraße 38, 70569 Stuttgart, Germany

July 10, 2019

All exercises are voluntary, for you to collect extra points.

1 Autoencoder (7 Points)

On the webpage find and download the Yale face database http://ipvs.informatik.uni-stuttgart.de/mlr/marc/teaching/data/yalefaces_cropBackground.tgz. The file contains gif images of 136 faces.

We want to compare two methods (Autoencoder vs PCA) to reduce the dimensionality of this dataset. This means that we want to create and train a neural network to find a lower-dimensional representation of our data. Recall the slides and exercises about dimensionality reduction, neural networks and especially Autoencoders (slide 06:10).

a) Create a neural network using tensorflow (or any other framework, e.g., keras) which takes the images as input, creates a lower-dimensional representation with dimensionality $p = 60$ in the hidden layer (i.e., a layer with 60 neurons) and outputs the reconstructed images. The loss function should compare the original image with the reconstructed one. After having trained the network, reconstruct all faces and display some examples. Report the reconstruction error $\sum_{i=1}^n \|x_i - x_i'\|^2$. (5P)

b) Use PCA to reduce the dimensionality of the dataset to $p = 60$ as well (e.g. use your code from exercise e08:02). Reconstruct all faces using PCA and display some examples. Report the reconstruction error $\sum_{i=1}^n \|x_i - x_i'\|^2$. Compare the reconstructions and the error from PCA with the results from the Autoencoder. Which one works better? (2P)

Extra) Repeat for various dimensions $p = 5, 10, 15, 20 \dots$

2 Special cases of CRFs (3 Points)

Slide 03:31 summarizes the core equations for CRFs.

a) Confirm the given equations for $\nabla_\beta Z(x, \beta)$ and $\nabla^2_\beta Z(x, \beta)$ (i.e., derive them from the definition of $Z(x, \beta)$). (1P)

b) Binary logistic regression is clearly a special case of CRFs. Sanity check that the gradient and Hessian given on slide 03:20 can alternatively be derived from the general expressions for $\nabla_\beta L(\beta)$ and $\nabla^2_\beta L(\beta)$ on slide 03:31. (The same is true for the multi-class case.) (1P)

c) Proof that ordinary ridge regression is a special case of CRFs if you choose the discriminative function $f(x, y) = -y^2 + 2y\phi(x)^\top\beta$. (1P)