TU Bergakademie Freiberg
Institut für Numerische Mathematik und Optimierung
Prof. Dr. O. Rheinbach
Dipl.-Math. Matthias Brändel

Wintersemester 2021/22
02.11.2021

Lineare Algebra, Datenanalyse und maschinelles Lernen 2

2. Exercise Sheet

Exercise 40

Write a `python` program. Load the `data_ex7.mat`. We will assume multiple different polynomial hypotheses, see (c).

(a) Write a function that calculates the value of the mean squared error cost function $J(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \mathbb{R}^d$,

$$J(\boldsymbol{\theta}) = \frac{1}{2}\frac{1}{m}\sum_{j=1}^{m}(\boldsymbol{\theta}^T \boldsymbol{x}_j - y_j)^2.$$

(b) Write a function that performs gradient descent on $J(\boldsymbol{\theta})$. Recall

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \alpha\frac{1}{m}\sum_{j=1}^{m}(\boldsymbol{\theta_i}^T \boldsymbol{x_j} - y_j)\boldsymbol{x_j},$$

where α is the learning rate or the step length of the algorithm.

(c) Assume three different hypotheses

$$h_{1,\theta}(x) = \theta_0 x,$$
$$h_{2,\boldsymbol{\theta}}(\boldsymbol{x}) = \theta_0 + \theta_1 x_1,$$
$$h_{3,\boldsymbol{\theta}}(\boldsymbol{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2^2.$$

Use the gradient descent algorithm to find the optimal $\boldsymbol{\theta}$ for each hypothesis. For $h_{2,\boldsymbol{\theta}}(\boldsymbol{x})$ use $\boldsymbol{\theta} = [4, 2]$ as starting parameters.

(d) For $h_{1,\theta}(x)$ and $h_{2,\boldsymbol{\theta}}(\boldsymbol{x})$ use a brute force approach to identify the optimal $\boldsymbol{\theta}$:

(d1) In $\mathbb{R}^1$ take the interval $\theta \in [0, 3]$, compute the cost function for sufficiently many values θ and plot the resulting curve. Plot the values of θ from the respective gradient descend history into the same plot and compare the results.

(d2) In $\mathbb{R}^2$ take the area $[-2, 4] \times [-2, 4]$, compute the cost function for sufficiently many values $\boldsymbol{\theta}$ (compare `matlab`'s `meshgrid` function for this approach) and plot the resulting cost values in a contour plot. Plot the values of $\boldsymbol{\theta}$ from the respective gradient descend history into the same plot and compare the results.

(e) Plot the data and the hypotheses in the same plot and discuss the results.