Mathematics of Learning – Worksheet 6 - Solution

- Solution sketches will be uploaded on June 07.
- You can hand in your own solutions and we correct them. You can either hand in by email (christian.biefel@fau.de) (until June 07) or on paper in any exercise class (until June 06). Please hand in in small groups of 2-3 students.
- For questions, please use the forum on StudOn or send me an email if it is a specific or personal question.

Exercise 1 [Regularization].
The generalized Tikhonov regularization is formulated as

$$RSS(\beta, \lambda) = \sum_{n=1}^{N} (y_n - x_n^T \beta)^2 + \lambda \sum_{i=1}^{M+1} \sum_{j=1}^{M+1} q_{ij} \beta_i \beta_j$$

for $\lambda \geq 0$ and with $q_{ij} = q_{ji}$.

1. Write $RSS(\beta, \lambda)$ in matrix notation.
2. Let $\lambda \geq 0$ be fixed. Compute the minimizer $\hat{\beta}$ assuming that the regularization term is convex. (What does that mean in the matrix notation?). State a condition under which it is unique.

Solution

1. $RSS(\beta, \lambda) = (y - X\beta)^T (y - X\beta) + \lambda \beta^T Q \beta$ with $X \in \mathbb{R}^{N \times (M+1)}$ and Q being the (symmetric) matrix with entries $q_{i,j}$.
2. The first derivative of RSS with respect to β has to be zero. This is sufficient, as RSS is convex under the assumption (Q positive semidefinite). We obtain the equation

$$-2X^T y + 2X^T X\beta + 2\lambda Q\beta = 0 \Leftrightarrow (X^T X + \lambda Q)\beta = X^T y.$$

   If $(X^T X + \lambda Q)$ is invertible (i.e., full rank), β is uniquely given by $\beta = (X^T X + \lambda Q)^{-1} X^T y$.

Exercise 2 [Examples].
Let $x = (1, 2, 3, 4, 5)^T$ and $y = (4, 2, 5, 7, 2)^T$.

1. Calculate $\beta_0, \beta_1, \beta_2, \beta_3 \in \mathbb{R}$ such that $\sum_{i=1}^{5} (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 - y_i)^2$ is minimal.
   Solution: $\min \sum_{i=1}^{5} (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 - y_i)^2 \iff \min \|X\beta - y\|^2$ with

$$X_i = (1, x_i, x_i^2, x_i^3)$$

and

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3).$$

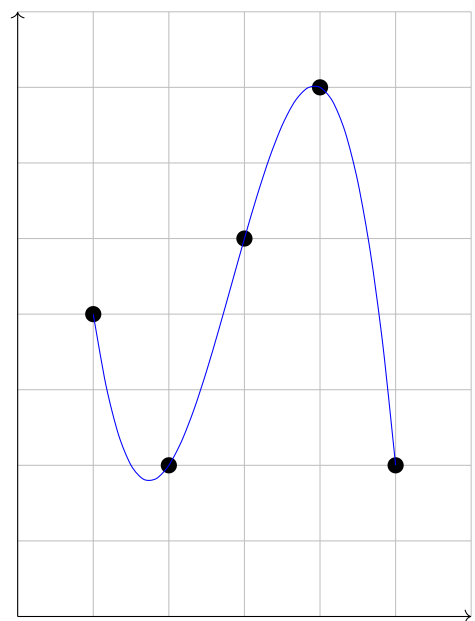
This is a linear regression problem, so everything we have to do is to solve the normal equations

$$X^T X\beta = X^T y,$$

i.e., with

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \\ 1 & 5 & 25 & 125 \end{pmatrix}$$

Solve the equations, we get the result: $\beta = (17, -\frac{41}{2}, \frac{17}{2}, -1)^T$, and if we plot the corresponding polynomial $17 - 20.5x + 8.5x^2 - x^3$, we get the following, fitting pretty well to our points:



2. Calculate $\beta_0, \beta_1, \beta_2, \beta_3 \in \mathbb{R}$ such that $\sum_{i=1}^{5}(\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 - y_i)^2 + 2 \cdot ||\beta||^2$ is minimal.
Solution. This is again a linear regression problem, this time with penalty parameter $\lambda = 2$. From the previous exercise we know that the optimal beta can be calculated using the formula

$$(X^T X + \lambda \mathbb{1})\beta = X^T y,$$

leading to the vector $\beta = (0.5662, 0.5103, 0.9120, -0.1857)^T$, and if we plot this, we get (fitting not that well; the gray plots are for some different penalty parameters $\lambda \in \{0, .01, .03, .3\}$)