FYS-STK4155 Week 35

Edvard B. Rørnes, Isak O. Rukan, Anton A. Brekke

August 28, 2024

Exercise 1

Show that

- $\frac{\partial(\mathbf{a}^T\mathbf{x})}{\partial\mathbf{x}} = \mathbf{a}^T$

  Using the summation notation where repeated indices are summed over, lower (upper) indices correspond to row (column) vectors and δ_{ij} as the Euclidean metric we have

$$\frac{\partial(a_i x^i)}{\partial x^j} = a_i\frac{\partial x^i}{\partial x^j} + x^i\frac{\partial a_i}{\partial x^j} = a_i\delta^i_j + 0 = a_j,$$

  which corresponds to $\mathbf{a}^T$ in matrix notation.

- $\frac{\partial(\mathbf{a}^T\mathbf{A}\mathbf{a})}{\partial\mathbf{a}} = \mathbf{a}^T(\mathbf{A}+\mathbf{A}^T)$

$$\begin{aligned}\frac{\partial(a_i A^i{}_j a^j)}{\partial a^k} &= a_i A^i{}_j\frac{\partial a^j}{\partial a^k} + \frac{\partial a_i}{\partial a^k}A^i{}_j a^j = a_i A^i{}_j\delta^j_k + \delta_{ik}A^i{}_j a^j \\ &= a_i A^i{}_k + A_{kj}a^j = a_i A^i{}_k + A_k{}^i a_i = a_i(A^i{}_k + A_k{}^i)\end{aligned}$$

  which corresponds to $\mathbf{a}^T(\mathbf{A}+\mathbf{A}^T)$ in matrix notation.

- $\frac{\partial(\mathbf{x}-\mathbf{A}\mathbf{s})^T(\mathbf{x}-\mathbf{A}\mathbf{s})}{\partial\mathbf{s}} = -2(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{A}$

$$\begin{aligned}\frac{\partial(x_i - s_j A_i{}^j)(x^i - A^i{}_k s^k)}{\partial s^l} &= -\frac{\partial s_j}{\partial s^l}A_i{}^j(x^i - A^i{}_k s^k) - (x_i - s_j A_i{}^j)A^i{}_k\frac{\partial s^k}{\partial s^l} \\ &= -\delta_{jl}A_i{}^j(x^i - A^i{}_k s^k) - (x_i - s_j A_i{}^j)A^i{}_k\delta^k_l \\ &= -A_{il}(x^i - A^i{}_j s^j) - (x_i - s_j A_i{}^j)A^i{}_l \\ &= -A^i{}_l(x_i - A_i{}^j s_j) - (x_i - s_j A_i{}^j)A^i{}_l \\ &= -2(x_i - A_i{}^j s_j)A^i{}_l\end{aligned}$$

  which is $-2(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{A}$. The next derivative is

$$\frac{\partial^2(x_i - s_j A_i{}^j)(x^i - A^i{}_k s^k)}{\partial s^l\partial s^o} = 2A_i{}^j A^i{}_l\delta_{oj} = 2A_{io}A^i{}_l$$

  which (if we differentiated w.r.t. $\mathbf{s}^T$ which I do not see why we don't? If this is the case just multiply by δ^{oj}) is $2\mathbf{A}^T\mathbf{A}$.

Exercise 2

The relatively low value for MSE suggests that our model's predictions are close to the actual values, and thus indicating a good fit. The value of R^2 is almost 1 suggesting that most of the variance in the data is accounted for. Increasing (decreasing) the coefficient in front of the noise term unsurprisingly suggests that we have a worse (better) fit to the data as there is more randomness involved.

Exercise 3

The optimal range for the polynomial degree seems to be around $10 - 12$ from running the code a few times. This is where the test data consistently has the lowest MSE suggesting that the model can be generalized to other cases. The lower end underfits the data resulting in a large bias. The MSE of the training data will of course be lower for higher polynomials, but will overfit the data as can be seen by the test data MSE diverging if one tries to run for, e.g. degree 30.