Faculty of Computer Science, Heinrich-Heine-University Düsseldorf

Deep Learning

Winter term 25/26 – Exercise Sheet 4

Submission Deadline: Monday, November 10, 2025, 2:00 PM

1. **Gradient**
   Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x^2 + y^2$. Sketch the graph of f in $\mathbb{R}^3$ by interpreting $f(x, y)$ as the height z at location (x, y). Calculate the gradient of f and draw some of those gradient vectors in the xy-plane. How does the height change if you move in the direction of the gradient?

2. **Chain rule**
   Let $g : \mathbb{R}^3 \to \mathbb{R}^2$ to be a multivariate vector function and $f : \mathbb{R}^2 \to \mathbb{R}$ to be a multivariate scalar function. They are given by $g(x_1, x_2, x_3) = (x_1^2 + x_2^3 + x_3 \quad x_1x_2x_3)$ and $f(y_1, y_2) = y_1^2 + y_2^2$.
   a) Calculate the gradient of f and Jakobian matrix of g.
   b) Calculate the gradient of the composition of $f \circ g$ with respect to x (i.e., gradient of function $f(g(x))$).

3. **Geometric interpretation of L2 regularization**
   a) Assume the loss landscape for a task at hand can be written down as

   $$\tilde{L}(\theta_1, \theta_2) = 0.25(\theta_1 - 4)^2 + 5(\theta_2 - 3)^2 \ .$$

   Minimize $\tilde{L}(\theta_1, \theta_2)$ and find $\hat{\boldsymbol{\theta}}$ in which it reaches minimum.
   b) Write down the Hessian matrix $\mathbf{H}$ of $\tilde{L}(\theta_1, \theta_2)$ and compute it at the optimal parameters $\hat{\boldsymbol{\theta}}$. Find its eigenvalues and eigenvectors.
   c) Recall from the lecture that if L2 regularization is added to the quadratic approximation of loss, then the parameters are decayed with respect to the eigenvalues of the Hessian

   $$\frac{\sigma_i}{\sigma_i + \lambda}$$

where λ is regularization parameter and σ_i is an eigenvalue corresponding to an eigenvector which is aligned with the considered parameter. Write down how each of the parameters is changing with $\lambda \in \{0.1, 1, 10\}$ using eigenvalues and eigenvectors of $\mathbf{H}$ computed before. What do you observe? Make a visual sketch of the effect of regularization in this example.