# Nonlinear Optimisation and Learning

*Resit* January 08, 2026        8:45 - 10:45 (two hours).

*No additional materials may be used during this exam (no notes, calculators, etc.). In your proofs, you may use definitions and all theorems, lemmata, corollaries and propositions from the reader with a reference like "We know that..." (make clear that you use a result from the reader). You may not use results from exercises. You may take this exam after handing in your work.*

This exam has 6 exercises.

1. Let $f(x,y) = (x-y)^2 + (y-1)^2$ on the domain $\mathbb{R}^2$.

   (a) Find the Hessian of $f$. Use the Hessian to show that $f$ is convex.

   (b) Find all critical points of $f$. Which of them are local minimizers? Justify your answer.

   (c) Find all global minimizers of $f$. Justify your answer.

   (d) Compute one step of Gradient Descent with Exact Line Search for $f$, starting from the point $(0,0)$.

   (e) Suppose we want to run two steps of Conjugate Gradient Descent on $f$, starting from the point $(0,0)$. What is the descent direction for the first step of CGD? What is the descent direction for the second step of CGD?

2. Let the function

$$f(x,y) = \tfrac{1}{2}x^4 + 2xy + 2x + (1+y)^2.$$

   (a) Suppose we want to do one step of Newton Descent for $f$, starting from the point $(0,0)$. What is the descent direction given by Newton Descent method?

   (b) Is the above direction a valid direction of descent? Justify your answer.

   (c) Determine the critical points of $f$, i.e., the points where $\nabla f$ vanishes.

   ✳ (d) Which of these critical points are local minimizers of $f$? Justify your answer.

   **NOTE**: *You may use the following fact about an invertible matrix $M \in \mathbb{R}^{2\times2}$.*

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad\Longrightarrow\quad M^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Define $h : \mathbb{R}^n \to \mathbb{R}$ as $h(\boldsymbol{x}) = f(A\boldsymbol{x})$. It can be shown that the gradient and the Hessian of $h$ are given by:

$$\nabla h(\boldsymbol{x}) = A^T \nabla f(A\boldsymbol{x})$$
$$\nabla^2 h(\boldsymbol{x}) = A^T (\nabla^2 f(A\boldsymbol{x}))A$$

   Show the following: If the function $f$ is convex, then the function $h$ must also be convex.

4. Several algorithms are used to train neural networks.

   a) Describe how the RMSProp differs from Adam.

   b) Is a neural network guaranteed to converge when trained with SGD? Briefly explain your answer.

   c) What is meant with "dropping the last batch"? Briefly explain your answer.

5. Neural networks can under- and overfit. Explain what is meant by these terms, and describe for both issues a way to prevent or resolve it.

6. Consider the function $f_\theta$ given by

$$f_\theta(x) = \text{ReLU}(w_1 x + b_1) + \text{ReLU}(w_2 x + b_2),$$

where $\theta = (w_1, w_2, b_1, b_2)$. We want to use gradient descent to fit the parameters. Our dataset consists of the single data point $(x, y) = (1, -1)$.

   (a) Write down the MSE loss function for this problem.

   (b) Sketch the computational graph for the loss function.

   (c) Compute the gradient of the loss function w.r.t. the parameters using backpropagation through the computational graph. Give the answer in terms of arbitrary parameter values for $(w_1, w_2, b_1, b_2)$. Your answer may contain ReLU and Step.

You can use that

$$\text{ReLU}(x) = \max(0, x).$$

and the derivative of ReLU is

$$\frac{d\,\text{ReLU}}{dx}(x) = \text{Step}(x) = \begin{cases} 1 & x \geq 0, \\ 0 & x < 0. \end{cases}$$

**Points:  90 + 10 = 100**

| | | | |
|---|---|---|---|
| 1. | (a) | : | 3 pt. |
| | (b) | : | 3 pt. |
| | (c) | : | 2 pt. |
| | (d) | : | 9 pt. |
| | (e) | : | 9 pt. |
| 2. | (a) | : | 6 pt. |
| | (b) | : | 6 pt. |
| | (c) | : | 6 pt. |
| | (d) | : | 9 pt. |
| 3. | : | 7 pt. | |

| | | | |
|---|---|---|---|
| 5. | (a) | : | 3 pt. |
| | (b) | : | 4 pt. |
| | (c) | : | 4 pt. |
| 6. | | : | 7 pt. |
| 7. | (a) | : | 2 pt. |
| | (b) | : | 5 pt. |
| | (c) | : | 5 pt. |