teachingML @ UoC

j.szymanski@fz-juelich.de

June 2025

Exam — Foundations of Machine Learning & Deep Learning

1. **Linear-Regression Gradient**
   (a) Given $\hat{y} = w_0 + w_1 x_1 + w_2 x_2$ and the loss $L = 12(\hat{y} - y)^2$, derive $\dfrac{\partial L}{\partial w_k}$ for $k = 0, 1, 2$.
   (b) Write the stochastic-gradient update for w_1 with learning rate η.

2. **Gradient-Descent Variants**
   Give two *advantages* and two *disadvantages* of each strategy:
   (a) Stochastic GD
   (b) Batch GD
   (c) Mini-batch GD

3. **Learning-Rate Tuning (Conceptual)** List two symptoms of using (a) too small and (b) too large a learning-rate when training a neural network.

4. **Underfitting vs. Overfitting** Describe one visual sign in the training/validation loss curves that indicates each condition.

5. **Bias and Variance (Intuition)** In one or two sentences each, state what "high bias" and "high variance" mean for a model's predictions.

6. **Weight Initialisation (Conceptual)** Why is random weight initialisation important for deep networks? Name two common schemes of weight initialisation (no formulas required).

7. **L2 (Ridge) Regularisation** Write the loss with an L2 penalty term for weights $\mathbf{w}$ and explain—in words—how increasing λ affects model complexity.
$$L = \sum_i \big(\hat{y}_i - y_i\big)^2 \; + \; \lambda \, \mathbf{w}_2^2$$

8. **Convolution Basics** An image $64\times64\times3$ is convolved with 32 kernels of size 5×5, stride 2, padding 1.

   Compute the output spatial size $H_{out} \times W_{out}$. Calculate the number of learnable parameters (include one bias per kernel).

9. **CNN vs. Fully-Connected on Images** Give two advantages a convolutional network has over a fully connected network when the input is an image.