First exam Machine Learning I – WiSe24/25, Kobak

90 minutes, no calculator nor sheet

4 points per task

---

1 – 3) Consider a linear model with 500 predictors. We fit our data with 1000 samples.

- Sketch test and train error of the ridge regression. The regression hyperparameter λ should be on the x-axis and the error on the y-axis. What measure of predictive performance makes sense? Give the mathematical definition of the measure used in the sketch.
- What happens if λ -> 0? What happens for λ -> ∞? Explain why and give the formulas necessary for your explanation.
- Now we remove 50 predictors for a fixed λ. How does the error change and why?

4) Let $y = \beta x$ be our model and $(x_i, y_i) = \{(2,1), (3,3)\}$ the data. What is the β that minimizes the loss?

5) For a linear regression model, we want to use the invented regularization term $||\beta-\beta_0||^{10}$ with β_0 as predefined constant. Write down the complete loss function $\mathcal{L}$ and the gradient $\nabla\mathcal{L}$.

6) Let X be a random variable. We draw ten independent samples $\{X_i\}^{10}{}_{i=1}$. Prove that the sample mean is an unbiased estimator of |E[X].

7) Consider logistic regression with two predictors x_1, x_2. The decision boundary goes through the points (1, 0) and (0, 2). Calculate the equation for the boundary and find all possible $\beta = (\beta_0, \beta_1, \beta_2)$ that lead to this boundary.

8) A convolutional neural network takes as input 100x100 RGB images. One layer consists of a 10x10 filter kernel with three feature channels. The network has 100 convolutional layers (with zero padding such that all layers have the same size) and one logistic regression output neuron.

A MLP takes the same input in a fully connected layer, a single hidden layer of the same size as the input and ends with one logistic output neuron.

Neglect the bias term for all layers for simplicity and compute the ration of parameters for both networks.

9) Is there a situation where the optimal k-means solution for k=2 has the same loss as the optimal solution for k=3? Provide an example dataset with more than two samples.

10) Construct a 3x3 covariance matrix in a way that the variance of variable 1 is 1, the variance of variable 2 is 0, the trace of the matrix is 5 and the covariance between variable 1 and 3 is 0.75.

11) Calculate all possible 2x2 covariance matrices where the first PC is pointing in the direction (1,1) and explains 100% of the variance.

12) If the data X was multiplied by 5, how would it affect the two-dimensional PCA projection and the tSNE embedding? Why?