October 7, 2025

Faculty of Computer Science, Heinrich-Heine-University Düsseldorf

Deep Learning

Winter term 25/26 – Exercise Sheet **ML Recap**

$\mathcal{X}$	Space of inputs (e.g., $\mathbb{R}^d$ or space of images)
$x \in \mathcal{X}$	sample from $\mathcal{X}$
$\mathcal{Y}$	Space of outputs/targets/labels (e.g., $\mathbb{R}$ for regression or $\{0, 1\}$ for binary classification)
$y \in \mathcal{Y}$	true label/target value in $\mathcal{Y}$
$\mathbb{P}_{\text{data}}$	data generating distribution over $\mathcal{X} \times \mathcal{Y}$
$\mathcal{H}$	hypothesis space, set of learnable models f
$f : \mathcal{X} \to \mathcal{Y}$	model/hypothesis, $f \in \mathcal{H}$
$f(x)$	output/prediction for input $x \in \mathcal{X}$
$L\big(y, f(x)\big)$	Loss function for sample $(x, y) \in \mathcal{X} \times \mathcal{Y}$
$\mathcal{D}_{\text{train}}$	Set of training samples $\big\{(x_1, y_1), \ldots, \{x_n, y_n)\big\}$, used for training a model
$\mathcal{D}_{\text{val}}$	Set of validation samples $\big\{(x_1, y_1), \ldots, \{x_n, y_n)\big\}$, used for fine tuning parameters for example
$\mathcal{D}_{\text{test}}$	Set of test samples $\big\{(x_1, y_1), \ldots, \{x_n, y_n)\big\}$, used for estimating the risk of the model, should not be used in any other operations with the model
$R(f)$	risk of model f, defined as $R(f) = \mathbb{E}_{(x,y)\sim\mathbb{P}_{\text{data}}}\big[L\big(y, f(x)\big)\big]$
$R_{\text{emp}}(f)$	empirical risk of f, for $\mathcal{D}_{\text{train}}$ defined as $R_{\text{emp}}(f) = \frac{1}{n}\sum_{i=1}^{n}\big[L\big(y_i, f(x_i)\big)\big]$

Note that we always consider multi-dimensional samples to be columns, e.g., having dimensionality $(p, 1)$, where p is the amount of features. Therefore, applying a model to

an input to get a prediction is always $\theta^T x$, thus parameters θ of the model are also a column (or multiple columns when output is multidimensional) of dimensionality $(p, 1)$. **Note** we consider models $f(x)$ that are characterized by parameters θ. The notation for such models is $f(\theta; x)$; sometimes we also use $f(x|\theta)$ to emphasize the dependency on parameters.

1. **Goal of Machine Learning**
   - Construction of a learner, i.e., an algorithm that learns a function (model/ hypothesis) from feature space to target space, based on training data, that maps unseen test data into label space proficiently.

2. **Components of a Learning Algorithm**
   - Hypothesis space (Space of functions "learnable" by our algorithm)
   - Evaluation (Performance measure of any given hypothesis in mapping unseen features to targets)
   - Optimization (Search method within hypothesis space)

3. **Different Kinds of Learning**
   - Supervised learning
   - Unsupervised learning
   - Reinforcement learning