## Practical Session 3

# Linear Regression

## 1 Weight regularization

**Problem 1:** Derive the closed form solution for ridge regression error function

$$E_{\text{ridge}}(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{w}^T\boldsymbol{\phi}(\boldsymbol{x}_i) - y_i)^2 + \frac{\lambda}{2}\boldsymbol{w}^T\boldsymbol{w}$$

Additionally, discuss the scenario when the number of training samples $N$ is smaller than the number of basis functions $M$. What computational issues arise in this case? How does regularization address them?

**Problem 2:** See Jupyter notebook `practical_03_notebook.ipynb`.

**Problem 3:** Using singular value decomposition of the design matrix $\boldsymbol{\Phi} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ show that predicted target $\hat{\boldsymbol{y}}$ for the training set when using $\boldsymbol{w}^*_{\text{ridge}}$ can be written as

$$\hat{\boldsymbol{y}} := \boldsymbol{\Phi}\boldsymbol{w}^*_{\text{ridge}} = \sum_{j=1}^{M}\left(\frac{\sigma_j^2}{\sigma_j^2 + \lambda}\boldsymbol{u}_j\boldsymbol{u}_j^T\right)\boldsymbol{y}$$

where $\boldsymbol{u}_j$ are the columns of $\boldsymbol{U}$, $d_j$ the elements of diagonal matrix $\boldsymbol{S}$ and $\lambda$ the strength of the $L_2$ regularization. What is the interpretation of this formula?

## 2 Multi-output linear regression

**Problem 4:** In class, we only considered functions of the form $f : \mathbb{R}^n \to \mathbb{R}$. What about the general case of $f : \mathbb{R}^n \to \mathbb{R}^m$? For linear regression with multiple outputs, write down the loglikelihood formulation and derive the MLE of the parameters.