Machine Learning	Prof. Matthias Hein
Exercise Sheet 11 - 17.6.2021 (16 Points)	due: 24.6.2021

Exercise 11 - Feature map for monomials

Let $x, y \in \mathbb{R}^n$ and $d \in \mathbb{N}$.

a. **(3 Points)** Show that the map defined componentwise by

$$\Phi_m(x) = \sqrt{\frac{d!}{\prod_{i=1}^n m_i!}} \prod_{i=1}^n x_i^{m_i}, \quad m \in M, \quad M = \left\{ \mu \in \mathbb{N}^n : \sum_{i=1}^n \mu_i = d \right\},$$

is a valid feature map for the reproducing kernel Hilbert space $\mathcal{H}$ with reproducing kernel $k(x, y) = \langle x, y \rangle^d$. This means you have to show that for all $x, y \in \mathbb{R}^n$

$$\langle \Phi_m(x), \Phi_m(y) \rangle = k(x, y) = \langle x, y \rangle^d .$$

b. **(2 Points)** Show that the dimension of $\mathcal{H}$ equals $\binom{d+n-1}{d}$.

c. **(1 Point)** Assume that the inputs x, y are images with 16×16 pixels and that $d = 5$. Compute the dimension of $\Phi_m(x)$, $\Phi_m(y)$. What are the problems of an explicit feature representation in this case ? Can the use of the kernel k provide remedy in this situation ?

Hint:

- When proving part a., you might use the Multinomial Theorem.
- When using induction in b., the identity $\sum_{k=0}^r \binom{k+l}{k} = \binom{l+r+1}{r}$ for non-negative integers l, r might be useful.

Exercise 12 - Consistency of model selection

(4 Points) Prove the following result for the minimization of risk of f' with the validation set of the size m.

Theorem 1 *Let f' be the classifier which minimizes the validation error $\hat{R}_m(f)$ among a set of classifiers in $\mathcal{F}$ with $|\mathcal{F}| = N$. Then with probability $1 - \delta$, we have*

$$R(f') \leq \inf_{f \in \mathcal{F}} R(f) + \sqrt{\frac{2}{m} \log\left(\frac{2N}{\delta}\right)}.$$

Exercise 13 - Cross Validation

Your task in this exercise is to find an optimal set of parameters for a classification problem using cross validation. The file `diabetes_data.npy` contains medical data from a study about the onset of diabetes in a high risk population of Pima Indians. The features correspond to certain physiological attributes and the class variables is +1 if a person shows signs of diabetes and -1 if not.

As a classifier use kernel ridge regression, i.e. the estimated function has the form $f(x) = \operatorname{sign}\left(\sum_{i=1}^n \alpha_i k(x_i, x)\right)$, where the coefficients α are computed by solving

$$(K + n\lambda \mathbf{1})\alpha = Y \ .$$

As kernel function, use the Gaussian kernel

$$k(x, x') = \exp(-\mu\|x - x'\|^2).$$

You are supposed to find the regularization parameter λ and the parameter μ using cross validation. As error measure use the 0-1-loss,

$$L(Y_i, f(X_i)) = \frac{1}{2}|Y_i - \text{sign}(f(X_i))|.$$

a. **(6 Points)** Implement in `CrossValidation` and apply 5-fold cross-validation on the training data `(Xtrain,Ytrain)` to determine the best parameters $\lambda, \mu \in \{10^{-4}, 10^{-3}, ..., 1\}$. Store the cross-validation errors for all possible values of (λ, μ) in the matrix `CVErrors` (with the convention that on the row one varies λ and on the column μ). Then train a new classifier on the full training set using the best parameters found and evaluate it on the testing set `(Xtest,Ytest)`. Report the parameters yielding the best cross validation error, as well as the obtained errors.

   Also compute training and test errors (of classifiers trained on the full training set) for all possible values of (λ, μ) and store them in the matrices `TrainErrors` and `TestErrors`. Create three 3D plots: training/test/cross-validation errors versus λ and μ (use a log-scale for λ and μ). Discuss the influence of λ and μ on all errors. How does the cross-validation error behave in comparison to the test error for different values of λ and μ?

Hints:

a. The surface plot can be done as described at `https://matplotlib.org/stable/gallery/mplot3d/surface3d.html`. Note that to get the log scale you can use the $x = \log(\lambda)$ and $y = \log(\mu)$.