QUIZ #2

1 - When using Regularized Linear Least Squares, we set the space of possible solutions to be

a. The space of linear functions $f(x)=x^Tw$
b. The space of any possible class of functions
c. The space of exponential functions of form $f(x)=e^{ax}$

we assume here the model associated with the function $\hat{f}$ we want to estimate is linear, $\hat{f}(x)=x^Tw$, $x,w \in \mathbb{R}^d$

2 - When the function f to be estimated from the data is non linear, we can resort to

a. Using a higher amount of input samples
b. A feature map, that maps the input to a new space where polynomial features are computed
c. A feature map, that maps the input to a new space (the feature space) where linear models can be employed

with the feature maps we use a trick and instead of working in the original space (where the model is non linear) we work in the feature space, where the problem is linear: $\hat{f}(x)=w^T\Phi(x)$

3 - A possible drawback in the use of feature maps (with original space of size d and feature space of size p)

a. Refers to the fact they can not be applied to classification problems
b. Refers to the computational cost, since in general p >> d
c. Refers to the representative power, as in general p << d

4 - In Regularized Linear Least Squares the role of the regularization parameter *lambda*

a. Is similar to the role of the K in K-Nearest Neighbours
b. Is to control the number of training samples
c. Is to control the number of points to be used for estimating the solution

The role of λ is similar to the role of K in KNN
a for small values of λ you tend to favour data fitting, for higher values of λ stability increases (at some point to the price of loosing the capability of representing the available data)

5 -A function that implements V-Fold Cross Validation to select the value of a parameter P in the case with V=2 has the following Input-Output parameters:

a. **Input**: Training Set (X,Y), proportion of training and validation data, learning method, range of parameters; **Output**: best error on the training set

b. **Input**: Training Set (X,Y), proportion of training and validation data, learning method, range of parameters; **Output**: best value for P, the one minimizing the validation error

c. **Input**: Training Set (X,Y), proportion of training and validation data, learning method, range of parameters; **Output**: best value for P, the one minimizing the training error

6 - A main difference between K-Nearest Neighbors and Regularized Linear Least Squares is that

a. The first is a global method, the second a local method

b. The first is a local method, the second a global method

c. They are both local methods but depend on different parameters

KNN is a local method as the output of a certain x only depends on the closest points. In RLS the output instead depends on all points of the training set.

7 - Solving Regularized Linear Least Squares amounts to

a. Solving a linear system

b. Solving a non-linear system

c. Solving an system of integrals

we start from the original problem

$$\hat{w}_\lambda = \underset{w \in \mathbb{R}^d}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i)^2 + \lambda \|w\|^2$$

and derived

$$(\hat{X}^T \hat{X} + \lambda n I) w = \hat{X}^T \hat{Y}$$

i.e. a linear system

8 - The effect of the regularization parameter (we called *lambda*) on the solution of Regularized Linear Least Squares is

a. To increase stability if small, to improve data fitting if large

b. To increase stability if large, to attenuate the effect of noise in the data if small

c. To increase stability if large, to improve data fitting is small

Having a look at the functional

$$\hat{w}_\lambda = \underset{w \in \mathbb{R}^d}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i)^2 + \lambda \|w\|^2$$

- when $\lambda \to 0$ also $\lambda \|w\|^2 \to 0$ thus stability is not preserved
- when λ increases the regularization term tends to have higher importance, improving stability
- when λ is too big, the regularization term wins on the data term, and the model looses memory of the training data (poor data fitting)

9 - Given a feature map Phi:R^d --> R^p, a main difference between the solution w of the Regularized Linear Least Squares in the original data space with respect to the one used when applying Phi is

a. In the size, d in the original space and 5d when using Phi

b. In the size, p in the original space and d when using Phi

c. In the size, d in the original space and p when using Phi

The functionals are the following:

$$\hat{w}_\lambda = \arg\min_{w \in \mathbb{R}^d} \frac{1}{m} \sum_{i=1}^{m} (y_i - w^T x_i)^2 + \lambda \|w\|^2$$ when working in the original space

$$\hat{\beta}_\lambda = \arg\min_{\beta \in \mathbb{R}^p} \frac{1}{m} \sum_{i=1}^{m} (y_i - \beta^T \Phi(x_i))^2 + \lambda \|w\|^2$$ when working in the feature space

10 - When using feature maps, selecting the appropriate type and size p of the feature space

a. Is not simple but a usual choice is p=10

b. Is not straightforward and a usual choice is to consider p very large (--> inf) and this corresponds to the use of Kernel functions

c. Is very simple: you can try them all