# TP 5 – AOS1
# Regularization

In the following, we will study the differences between the classical linear regression model and their most popular regularized versions: the ridge regression model and the Lasso model.

We will use the sklearn Python library and some tools from the scipy library.

```
from sklearn.linear_model import Ridge, LinearRegression
from scipy.linalg import svd
```

A custom class is provided in the file utils.py that models a linear random phenomena. You have to first create the random phenomena

```
from utils import Event
evt = Event(n_features=n_features, effective_rank=3, noise_level=1)
```

Then you can repeatedly sample from it

```
batch1_X, batch1_y = evt.sample(n_samples=100)
batch2_X, batch2_y = evt.sample(n_samples=200)
```

More formally the following model is implemented inside Event

$$y = X\beta + \varepsilon.$$

The coefficients are randomly generated unless given as argument to `Event`. They are accessible with `evt.coefficients`.

The matrix $X$ is created on demand when getting samples with the `sample` method. It is controlled by the argument `n_features` (10 features by default) and the `effective_rank` integer that is the "denoised rank" of $X$.

The standard deviation $\sigma$ of the noise $\varepsilon$ is controlled by the noise_level parameter in Event. More precisely, noise_level indicates a ratio in percentage between $\sigma$ and $\|\boldsymbol{\beta}\|$. The standard deviation is available in the attribute model_noise.

1. Using a SVD, compute the singular values $d_i$, $1 \leqslant i \leqslant k$ of $X$. What is the effect of the parameter `effective_rank`? What is the effective rank supposed to model ?

2. By repeatedly fitting each model on the same Event object, give an estimate of the bias, variance and risk of the estimator at a chosen point.

3. Show the influence of the dataset's effective rank on the 3 algorithms and interpret.

# Problem: make elastic net outshine the lasso

The elastic net regression was invented to compensate for the lack of robustness of the lasso regression. The elastic net especially outshines the lasso when some variables are highly correlated and on the same scale. You have been shown in the last course's slides some intuitive arguments demonstrating why it should particularly be more robust in this case. But could you demonstrate this experimentally?

Design a regression dataset in which we want to select variables and where the elastic net gives better results in terms of stability of the set of selected variables.