HW2 - Regression Regularization

Ridge regression

Recall that for linear models with scalar output we have $h(x) =< w, x >$ and that the Empirical error $L_S(h)$ can be written
in terms of the vector of parameters w, in the form:

$$L_S(w) = \frac{1}{m_t} \|Y - Xw\|^2 \tag{1}$$

where Y and X are the matrices whose $i-$th rows are, respectively, the output data y_i and the input vectors $x_i^\top$.

In the case of Ridge regression we add a regularization term to the RSS term so that our Empirical error becomes:

$$L_S(w) = \frac{1}{m_t} \|Y - Xw\|^2 + \lambda \|w\|^2 \propto \|Y - Xw\|^2 + \underbrace{\lambda * m_t}_{:=\alpha} \|w\|^2 \tag{2}$$

The Ridge Least Squares solution is given by the expression:

$$\hat{w}_{Ridge} = \arg\min_w L_S(w) = (X^\top X + \alpha I)^{-1} X^\top Y \tag{3}$$

Note: what has changed w.r.t. the LS solution? Do we need to worry about invertibility of the matrix we need to invert?

- Prove that adding a positive multiple of identity to a semi definite positive matrix you get a positive definite matrix.
- Prove that a positive definite matrix is *always* invertible.

LASSO

A different regularization to our linear model: LASSO - Least Absolute Shrinkage and Selection Operator (l1 regularization)

Note: the parameter λ is called α in the Lasso model from sklearn.

Homoschedasticity vs Eteroschedasticity

Eteroschedasticity (variance of the regression variable depends on the independent variable).

Homoschedasiticty (variance independent on the independent variable).