# Machine Learning for Physicists - First Assignment

Guilherme Simplício

October 31, 2022

## 1 Theory Part

1)You are training a ridge regression model with zero regularization. How does the training loss behave for n < d? Explain your answer.

> If we are training ridge regression model with $\lambda = 0$, we are in the domain of linear regression, where $\hat{w} = (X^T X)^{-1} X^T y$. In this case there is only a general solution for $\hat{w}$ if $X^T X$ exists, ie, if n≥ d.Thinking in terms of rank(or non trivial eigenvalues), they are a increasing function for n < d. For this reasons, if we plot the train error(loss) we would see that it would decrease until $\alpha = \frac{n}{d} = 1$

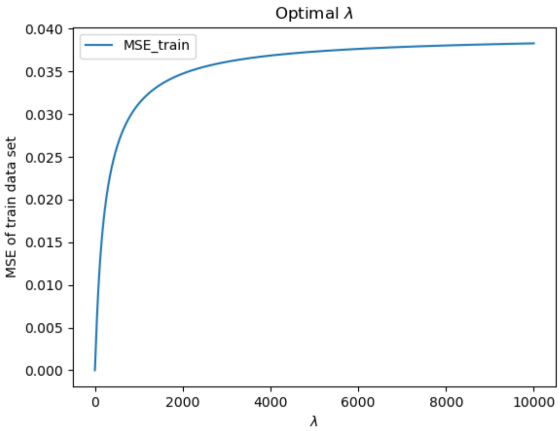
2)Let X ∈ $\mathbb{R}^{n \times d}$ be the training data matrix and y ∈ $\mathbb{R}^n$ be the labels. Under what condition on X and y is ridge regression with zero regularization able to exactly fit the training points?

> As it was explained on 1), there is only a general solution for $\hat{w}$, in the zero regularization regime, if there are at least as many parameters as equations (ie n ≥ d) provided they are linearly independent, this means that the matrix $X^T X$ is full rank, hence invertible. Furthermore, the augmented matrix [X|y] has to be the same rank as X, in order to avoid fits that the same X has two different y's.

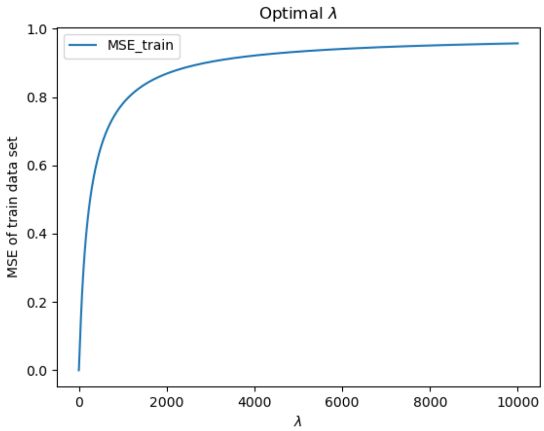
3)You train ridge regression several times with fixed n and increasing regularization $\lambda$. How do you expect the training error to depend on $\lambda$ ? Express the training error at $\lambda = \infty$ in terms of the labels of the training set y = $(y_1, y_2, ..., y_n)$. Explain your reasoning.

> Regarding the **training** error, as $\lambda \rightarrow \infty$, tending to a horizontal asymptote to the variance(=$\sigma^2 = \frac{\sum(y-\bar{y})^2}{n}$) of the training set y. As the regularization strength approaches high values, the training does not depend on the input, hence the the output tends to a mean value, this means that the mean squared error is actually the variance!

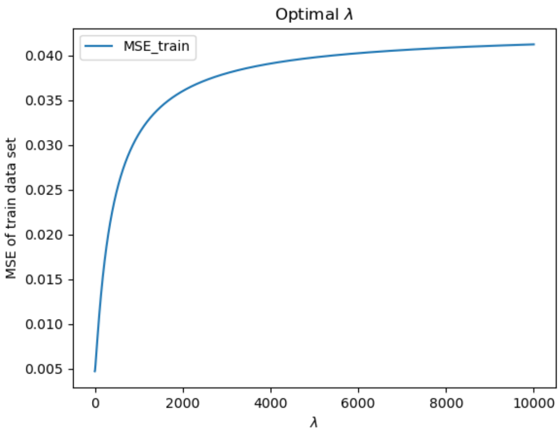
**Bonus for 3:** I slightly modified the code done in main.py and created main2.py (commited in Github) in order to produced this plots who illustrate the aforementioned behaviour.



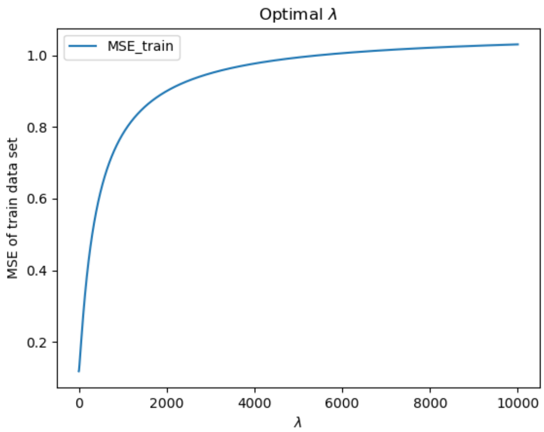

(a) $\sigma^2$=$(0.2)^2$ = 0.04, n=80.




(b) $\sigma^2$=$(1.0)^2$ = 1.0, n=80.




(c) $\sigma^2$=$(0.2)^2$ = 0.04, n=150.




(d) $\sigma^2$=$(1.0)^2$ = 1.0, n=150.

Figure 1: Optimal $\lambda$ for Ridge Regression, considering training data set.

## 2 Acknowledgments

I would like to thank Francisco Simões and Arianna Alonso Bizzi for having very useful debates with me regarding this assignment and the course, in general.