Lecture Notes : Basics of ML

Scott Pesme
INRIA Grenoble

September 29, 2025

Contents

1 Introduction

Slides of the general introduction can be found here.

1.1 Regression

Your content here.

A Cheat Sheet of Useful Formulas

A.1 Definitions and notations

Keep in mind that these are the notations I like to use, but these are obviously personal and others will use different ones!!

In the following I use the following notations:

- x will always correspond to data. E.g. $x_1, \dots, x_n \in \mathbb{R}^d$ could be some dataset. n is then the number of training samples, and d the dimension of each data point.
- $X = \begin{pmatrix} - & x_1 & - \\ & \vdots & \\ - & x_n & - \end{pmatrix} \in \mathbb{R}^{n \times d}$ corresponds to data / feature / design /observation matrix. It has $n =$ *"number of samples"* rows and $d =$ *"dimension of datapoints"* columns.

A.2 Gradients

If $f : \mathbb{R}^d \to \mathbb{R}$ is differentiable, then its gradient $\nabla f : \mathbb{R}^d \to \mathbb{R}^d$ is defined as: $\nabla f(w) = ...$

- For a vector $b \in \mathbb{R}^d$, the gradient of the linear function $f : \mathbb{R}^d \to \mathbb{R}$, $f(w) = \langle a, w \rangle$ is equal to $\nabla f(w) = b$.
- For a (not necessarily symmetric) matrix $A \in \mathbb{R}^{d \times d}$, $f(w) = w^\top A w$ is a quadratic function. Its gradient is $\nabla f(w) = \frac{1}{2}(A + A^\top)w$, which is equal to Aw iff A is a symmetric matrix. The hessian of f is equal to $\nabla^2 f(w) = \frac{1}{2}(A + A^\top)$.

A.3 Linear Algebra

- For a matrix $A \in \mathbb{R}^{d \times d}$, it holds that $w^\top A w = \sum_{i,j=1}^{d} w_i w_j A_{i,j}$.

A.4 Good practices and sanity checks