MATH 412 BASICS

Rafiki's Notes

Rafael Barroso
Mathematics Ing.
EPFL
September 14, 2025

Contents

1 Supervised Learning Basics

In this area of machine learning, we try to understand certain relations beween input-output data. If such relations are established, we then wish to generalize for new *unseen* data. Things start getting even jucier whenever we wish to take decisions based on new data, resulting in a more generalized task. In this section we explore the formalization provided by Prof. Obozinski.

1. **We have:**
    - Data: $\mathcal{D}_n := \{(x_0, y_0), \ldots, (x_n, y_n)\}$
    - i.e. tuples of the form (x_i, y_i)
    - $x_i :=$ input; $y_i :=$ output
2. **We want:**
    - Given $\mathcal{D}_n$, learn relations of the x_i's with the corresponding y_i's such that we may infer something about a new unseen y' given x'.

We now define the two types of tasks cosidered inside supervised learning (amongst others).

Definition 1.1. A *prediction* task is established to be the discovery of y‘ (unseen) given x‘. A *decision* task on the other hand, focuses on producing a decision based on (x‘, y‘) only with the data of x‘

For example, take into consideration a medical diagnosis. We have $x_i :=$ patient data i.g. $\{\text{weight}_i, \text{height}_i, \ldots\}$; $y_i := \{\text{positive}, \text{negative}\}$. Then, a **prediction task** would consist in predicting y‘ given x‘. A **decision task** on the other hand, would then consist on choosing how to treat patient x‘ i.g. choosing medicine $m \in \{A, B, C\}$ (we have to decide on y‘ by only seeing x‘).

We now consider the space of all possible decisions; a *learning algorithm* (sometimes called *learning scheme*) $\mathcal{A}$.

Definition 1.2. We define a learning algorithm as

$$\mathcal{A} : \mathcal{D}_n \to \hat{f}$$

where $\hat{f}$ is our decision function.

Obviously we want $\hat{f}$ to be "good" (otherwise, *nos estamos haciendo pendejos*). Hence, we must define what it means for $\hat{f}$ to be "good" i.e. what we want from $\hat{f}$.

Definition 1.3. Let $\mathcal{X}$ be the input space, then, a decision function is defined as

$$f : \mathcal{X} \to \mathcal{A}^{\mathcal{X}}$$

Note that the input space $\mathcal{X}$ is the space of all x_i's.

Ideally, as stated before, we want a "good" function (i.e. decision function) f such that $f(x) \in \mathcal{A}^{\mathcal{X}}$ is "good" when compared to an unseen y. This means that $f(x)$ must be an accurate prediction of y and it has the **smallest possible cost** whenever y occurrs. So, we compute the *loss function* l.

Definition 1.4. Let $\mathcal{Y}$ be the space of all possible outcomes, then

$$l : \mathcal{A}^{\mathcal{X}} \times \mathcal{Y} \to \mathbb{R}$$

defined by $(f(x) = a, y) \mapsto l(a, y)$. Note that this function meassures the cost of taking decision f whenever y occurrs i.e. the **risk**.

[1]

References

[1] D. S. Judson, *Abstract Algebra: Theory and Applications*, 3rd ed. Orthogonal Publishing L3C, 2019.