CS 412 — Introduction to Machine Learning (UIC)	January 23, 2025

Lecture

Lecturer: Aadirupa Saha *Scribe(s): Uwadia Edomwonyi*

Data Observations → Machine Learning System → f (prediction functions/prediction algorithms)

Primer on Functions:

- A function takes an input value, performs a specific set of operations on it based on a defined rule, and then produces a single, corresponding output value; essentially, it establishes a relationship between an input and an output, where each input is mapped to only one unique output.
- A function from a set X to a set Y assigns to each element of X exactly one element of Y. The set X is called the domain of the function and the set Y is called the codomain of the function.
- In machine learning, X is the input set and Y is the output/label set

Examples of prediction tasks:

- Rain Prediction
  X:{set of all possible weather conditions} x $\in X \subseteq \mathbb{R}^d$
  $Y : \{Rain, NoRain\} = \{0, 1\}$.

- Digit Recognition X:{finite set of images} $x \in X \subseteq \mathbb{R}^{307,200}$ (# of pixels in an image, each x has a grayscale value)
  Y:{0,1,2, ... ,9}.
- Predicting the Sale Price of a House X:{features of the house, zip code, square footage, age, etc.} $x \in X \subseteq \mathbb{R}^d$ (d = of features)
  Y: price of a house in dollars $\subseteq \mathbb{R}_{>0}$ [10 USD - 10,000,000 USD]
  Note that we can use One-Hot Encoding to encode categorical attributes

How to find predictor function, f given domain D

D= $(x_1, y_1), ..., (x_n, y_n)\}$ Dataset of size n, where $x \in X$ is an input and $y \in Y$ is an output. And the function, f maps all $x \in X$ to all $y \in Y$.
Let's suppose the following: ∃f*: X $\rightarrow Y s.t.$f$^*(x_i) = y_i, \forall i \in [n]$. If we do find such an f^* then we are good!

We use a loss function, l, to help us understand the error between the predicted and actual values.
In supervised machine learning a loss function, l, takes the true output space (the actual target value) and the predicted output space (the model's prediction) as inputs, and calculates a numerical value representing how much the prediction deviates from the true value, essentially measuring the "error" made by the model; the goal is to minimize this loss value during training to improve the model's accuracy
Y = Actual
Ŷ = Predicted
Error = Y - Ŷ

Depending on the task/problem, different loss functions are used to assess the quality of the model (i.e., the prediction function). We will discuss Regression and classification (binary classification and multi-classification):

Regression Loss Functions; Y $\subseteq \mathbb{R}, \hat{Y} \subseteq \mathbb{R}$
$Absoluteloss\, |y - \hat{y}|$
$Squaredloss(y - \hat{y})^2$
Regression Loss Functions; Y $\subseteq \{0,1\}^m, \hat{Y} \subseteq \{0,1\}^m, wheremistheofclasses$
$LogisticLoss(y, \hat{y}) = -y * log(\hat{y}) - (1 - y) * log(1 - \hat{y}), \forall y \in Y, \forall \hat{y} \in \hat{Y}$
$MulticlasslogisticLoss(y, \hat{y}) = \sum_{i=1}^{m} y_i * log(\hat{y}_i), \forall y \in Y, \forall \hat{y} \in \hat{Y}$