CMSE 820: Math Foundation of Data Science.

1 Supervised Learning:

learning the function between the input and output based on example input-output pairs.

Given training data:

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(n)}, y^{(n)})\},$$

Goal: build a model $y = f(x)$ to predict the corresponding y for new x.

Examples:

- y numerical: regression (e.g. linear, ridge, lasso)
- y categorical: classification (e.g. perceptron, linear discrimination. support vector machine.

2. Unsupervised learning: look for patterns in a data set with no pre-existing labels.

Given data

$\{ x^{(1)}, x^{(2)}, \cdots, x^{(n)} \}$

Goal: find hidden patterns in the data.

Examples:

- dimension reduction: (e.g. principal component analysis, multi-dimensional scaling)
- clustering

Notations:

- $\forall$: for all / any
- $\exists$: there exist(s)
- $\Rightarrow$: implies
- $\Leftrightarrow$: equivalent
- $\mathbb{R}$: real line
- $\mathbb{R}^n$: collection of n-tuples $(x_1, \cdots, x_n)$
- $\mathbb{S}^n \subset \mathbb{R}^{n+1}$: n-dim unit sphere.

Linear algebra:

Def: (Linear Dependence / Independence)

The vectors $x^{(1)}, x^{(2)}, \cdots, x^{(n)}$, are called <u>linearly independent</u> if

$$c_1 x^{(1)} + \cdots + c_n x^{(n)} = \vec{0}$$

implies $c_1 = \cdots = c_n = 0$.

Otherwise, these vectors are <u>linearly dependent</u>.

A matrix $A \in \mathbb{R}^{m \times n}$ is a diagram of numbers with m rows and n columns.

Def: $\mathrm{Ker}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$

$\mathrm{Ran}(A) = \{y \in \mathbb{R}^m : Ax = y$ for some $x \in \mathbb{R}^n \}$