20180116 COGS 118 Lecture Notes

Cabinet Cogs118a Lecture Notes

Supervised Classification

Three key variables

1. Input: $x = (x_1, x_2, \dots)$
2. Label: $y \in 0, 1$
3. Model Parameter: W

Vector Calculus

- Addition
- Scaling
- L2 Norm: $||a||_2 = \sqrt{\sum_{i=1}^{n} a_i^2}$, $||a||^2 = \sum_{i=1}^{n} a_i^2$
- L1 Norm (sparse): $||a||_1 = \sum_{i=1}^{n} |a_i|$
- Projection (inner product/dot product)
  - Note that it is a *scalar*
  - Important bc every sample lies in high dimensional space and the problem of **matching** is analogous to finding a projection of 0. Quantified similiarity.

- Decision boundary: $\{x_i, f(x_i) = 0\}$, $f(x) = a \times x_1 + b \times x_2 - c$
  - The boundary is a **set**.

  - Also, $\frac{a}{c}x_1 + \frac{b}{c}x_2 + 1 = 0$
  - Also, $\frac{a}{\sqrt{a^2+b^2}}x_1 + \frac{b}{\sqrt{a^2+b^2}}x_2 + \frac{c}{\sqrt{a^2+b^2}} = 0$
  - $f(x,w) =< w, x > + \frac{c}{||w||_2}$
  - $f(x) = 0 \Rightarrow \frac{f(x)}{c} = 0$
  - For linear transformation, there is a direction and a translation.
  - w is orthgonal (normal) to the boundary.

Distance to the decision boundary

Compute the projection of the normal (w) with the input (x):

$$w^T x =< w, x >$$

Now determine if $w^T x < 0$ or $w^T x > 0$

In higher dimenstions, this line becomes a **Hyper-plane**.

Vector Derivatives

- Vector-by-scalar
- Vector-by-vector
- Matrix-by-scalar
- Scalar-by-vector

Probability

- Deterministic
- Random
- Axioms of probability
- Independence of events
- Baye’s theorem