2020年9月16日 8:01

Training loss $L_{train} = L(f, X_{train}, Y_{train})$.

Test loss $L_{test} = L(f, X_{test}, Y_{test})$.

Validation loss $L_{valid} = L(f, X_{valid}, Y_{valid})$.

These pairs of (X, Y) are sampled from population loss D_X, D_Y.

↳ ultimate goal: minimize $_f$ $L_{population} = E_{X,Y \sim D_X, D_Y} L(f, X, Y)$.

Overfit vs. Underfit.

- Underfitting - your function may not have enough representation power
- overfitting - your function has too much representation power.
  - gives $L_{train} \simeq 0$, but generalizes bad.
  - Regularization.

Modern View:

For some neural network, there are implicit regularizations

↳ explicit regularization not neccessary.

(perhaps because of SGD algo.).

Unsupervised Learning

- Clustering ⟶ Data mining / recommendation system.
- Principal component Analysis (Find the most important components (directions)
  ↳ minimize MSE.
- Generative Model ⇒ Describe distribution of data.
  by mapping Gaussian to target distribution
- Anomaly Detection.
- Dimensional Reduction

Semi-Supervised Learning

data < labeled
       unlabeled

Can we use unlabeled data to improve prediction?

Assumptions: points with small distances
are more likely to share same labels.