Linear Classification
Support Vector Machines

Srinivasan Viswanathan

Nov 9, 2022

Recap

- Recap.
- Linear Classification problem
- Perceptron algorithm for decision boundaries
- Support vector machines
- Lab/Assignment discussion

Big picture again..

- Feature Vector $X \in R^d$
- Target/Labels $Y \in R$
- Training Set (X^i, Y^i)

- Hypothesis Function (Regression or Classification) - $h_\theta(X)$
- Cost Function (objective function, training error etc)

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

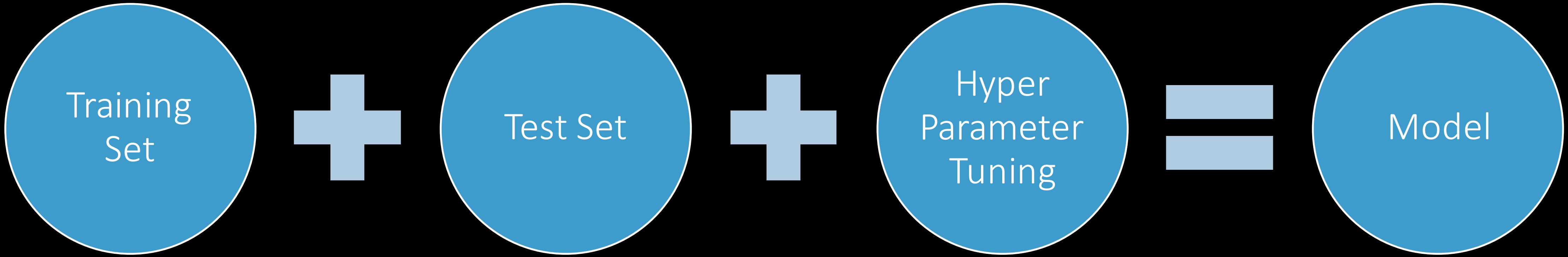
Big picture again (ctd)..

- Cost Function with Regularizer: (will vary)

$$J_\theta = \frac{1}{2m}\sum_{i=1}^{n}(h_\theta(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2} \parallel w \parallel^2$$

- Testing Set (X^i, Y^i)

Big Picture (Ctd)



$h_\theta(X)$

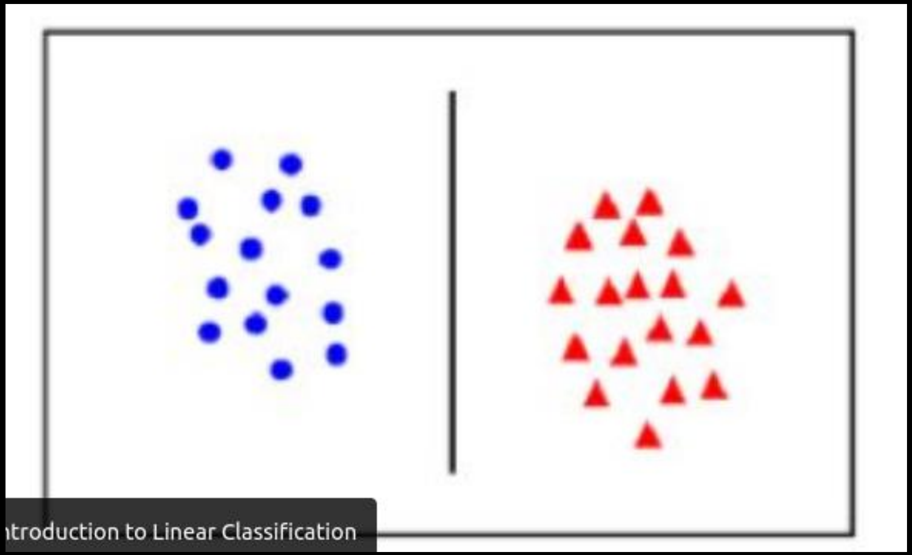
Testing Set or Validation Test + Hyper Parameters for regularizers together guides the hypothesis function to have generalization properties. Thus selecting best hypothesis function among the family.

Introduction to Linear Classification

- Is Movie good or Bad ? Ie. The label is –1 or +1.
- Feature vector for Movie – actor, language, story theme, type (horror, action etc.), audience type etc.,
- The hypothesis function will map the features to label –1 or +1.
- The hypothesis is a Linear function, (line in two dimension or hyperplane in n dimension)..

$$\theta_0 + \theta_1 x_0 + \theta_2 x_1 + \ldots + \theta_n x_n$$

$$\theta^T X + \theta_0$$



$$\theta^T X + \theta_0 = 0$$

Decision Boundary

$$\theta^T X + \theta_0 > 0$$

Label: +1

$$\theta^T X + \theta_0 > 0$$

Label: -1

Training Error ?

- Training Error /cost function sum of bad labeling

$$\frac{1}{n}\sum_{i=1}^{n}(h_\theta(x^i) \neq y^i)$$

- It could also be written as :

$$\frac{1}{n}\sum_{i=1}^{n}(y^i\theta x^i \leq 0)$$

Examples.. (interactive)

- Discuss examples of bad decision boundaries
- Examples of good decision boundaries
- Is it linearly separable (?)

Linearly Separable

- Training Samples are linearly separable if there exists parameter such that

$$y^{(i)}(\theta * x^{(i)} + \theta_0) > 0$$

for all $i = 1, \ldots, n$

Perceptron Algorithm

$$\textbf{if} \quad y^{(i)}(\theta \cdot x^{(i)}) \leq 0 \textbf{ then}$$
$$\theta = \theta + y^{(i)} x^{(i)}$$

Perceptron Algorithm

```
for iter in range(0,niterations):
    for i in range(0,N):
        if y[i]*(theta*x[i]) <= 0:
            theta = theta + y[i]*x[i]
            # offset separately handled below
            theta_0 = theta_0+y[i]
```

Note: algorithm above is pseudo code for discussion only. Theta, x y are vectors so needs to be handled properly.

Problems with any linear separator - Interactive

- Discuss examples of bad decision boundaries
- Examples of good decision boundaries
- Perceptron algorithm always Finds a solution if it is linearly separable.

Support Vector Machines

- Optimization Function with Hinge Loss and Regularizer.
- Find the Optimal Hyperplane that will separate the points.
- Leads to more generalization and better behavior for test/validation sets.
- Hinge LOSS:

Thank You !

Srinivasan Viswanathan