Lecture 20

Name: Harsh Sanjay Roniyar

Roll Number: 22B3942

Discussing some issues with Gradient descent, as understood in Lecture 19.

We use **Stochastic Gradient Descent** (SGD) to escape local minima and other critical points.

`SGD:`

```
loop{
        i <- random sample [N]     --- sgd

        l_i <- l(t_i, x_i, theta)  --- scalar
        grad_theta(l_i)            --- vector

        theta_new <- theta_old - eta * grad_theta(l_i)
}
until{
        stopping criterion
}
```

Batch Gradient Descent

`BGD:`

```
batch formation{
        Randomly permute training samples
        Divide training data into batches of size N/B samples
}

epoch loop{
        for each batch b in [B] {
                L_batch = sum(l_i) for i in batch
                grad_theta(L_batch) = (sum(grad_theta(l_i)) for i in batch)
/ (N/B)
                theta_new = theta_old - eta * grad_theta(L_batch)
        }
}
until{
        stopping criterion
}
```

Now, lets see how we can combine models

Combining Models

`ensemble` models

$\rightarrow$ for classification

- model averaging:
  - assumption 1: probability that any model is correct if p > 0.5
  - assumption 2: h_1, h_2, h_3 are independent of each other

nearest neighbor
splits into voronoi regions
label is the same as the data point