Lecture 2 | Linear Regression and Gradien Descent

Date de création	@December 14, 2024 5:18 PM
Étiquettes	

Notes

Gradient Descent

To apply gradient, we need:

- start with some random θ
- keep changing θ to minimize $J(\theta)$

Since we start θ randomly maybe you end in different places at the end (but still lower than the beginning).

We apply gradient until convergence. and in each interation you update for j = 0, 1, 2, 3…

Batch Gradient Descent

This gradient is called Batch Gradient Descent and it’s negative side is that in large dataset it becomes slower.

Stochastic Gradient Descent

```
Repear {

    for j = i to m {
        wi := wi - lr(woxi - yi)xji
    }
}
```

Supose we have 100 milion exemples, GD one step requires you run trought all data.

SGD updates at every interations (for every single element).

If you’re using Linear Regression Algorithm, It’s possible in one step go to the local minima (normal equation).

Quotes

Questions

▼ How compute PD **J(**θ)

PD = Partial Derivative

(hθx - y). PDθj(θoxo + θixi + .. - y), and we now that, for all values that aren’t θj the values of PDθj is zero.

So, the final result is (hθ(x) - y) * xj

▼ How is a step in gradient descent?

θj = θj - lr * sum (i=0, m) (hθ(x^{i}) - y^{i})xj^{i}

▼ How find the Global Minima for Linear Regression in one single step?

Normal Equation

$$X^T\Theta = X^Ty$$
$$\Theta = (X^T)^{-1}X^Ty$$