**basic part**

我這次用的演算法是式gradient descent，我以下會講解每個function的功能

1. SplitData:

我把data分成三個部分，第一筆有159筆資料當作train data，第二筆有30筆資料當作validation data，以便於測試我model的準確度，第三筆有20筆資料，預測10/15~11/11的資料。

1. PreprocessData

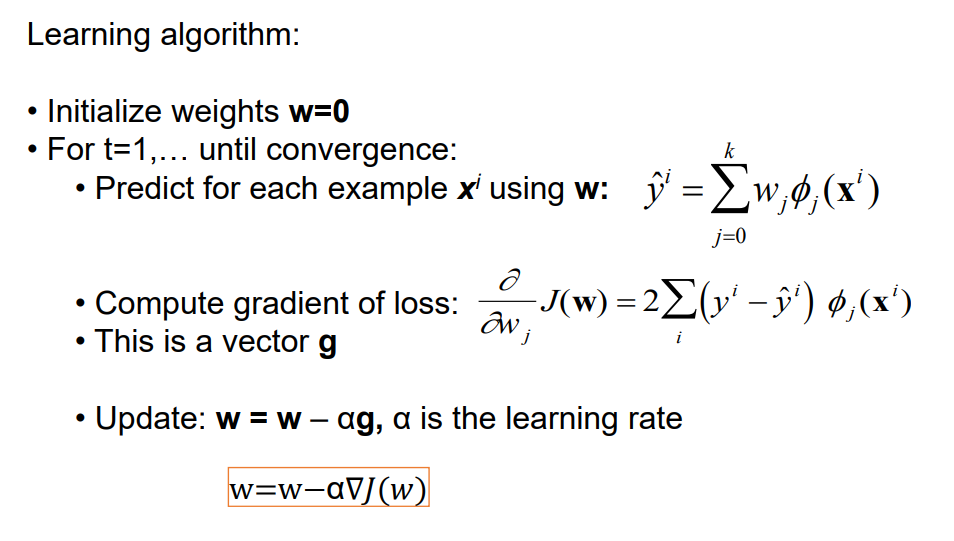
分別輸入分為兩個部分，MTK的data當作X，TSMC的data當作Y(ground truth)。因為我的f(x)=，所以我的X會變成陣列，變成[]。

1. Regression

核心概念就是w=w-learning\_rate\*g，w的直要不斷被更新，我的w的值一開始是設為[0,0,0,0,0]，而g的值會在CountLoss中被計算出來。而每個係數都有不同learning\_rate，經過不斷的測試 ，我將learning\_rate設為，而iteration設為20000次。當iteration跑到20000次的時候，w的值基本上已經收斂了。

1. CountLoss

這個function的作用就是計算g，用到的公式就是講義上的公式，如下圖



1. MakePrediction

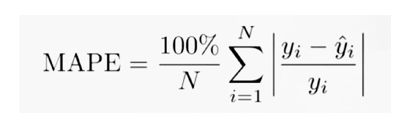
最後一步就是拿計算好的w，去乘上test data。我最後的w為

[-2.34478473e-11,-1.29854370e-07,-2.06041151e-04,

9.42738624e-01,2.15721864e+01]

以上就是我每個function的作用，以下會再補充說明幾點:

1. 我另外有寫了一個evaluate function以測試我的model的準確度，用的方法是ppt上寫的算法(如下圖)，我在train model時，也是看mape的值去判斷當前model的好壞。做測試的時候，我拿validation data去跑model，得到的值是2.59%。



**bonus:**

我這次用的演算法是式gradient descent與Autoregressive(AR)，除了MTK之外，我還有用UMC(聯電)今年的data，因為他的產業發展跟台積電蠻像的。有些部分跟basic很像，所以以下會針對不一樣的地方做講解。

做法主要分兩個部分，第一個是用AR的方式，分別預測MTK跟UMC在10/15~11/11這20天內的價錢。第二個部分則是運用basic的方法，用1/1到10/14的data，找出MTK跟UMC與TSMC的相關性(w)，再將這個w套用到之前MTK跟UMC預測出來的資料，由此預測10/15~11/11 TSMC的data。以下會分成這兩個部分做講解:

**第一部分:**

我以MTK做講解(UMC的做法是一樣的)，我的train data是取MTK前四天的data去預測今天的data，而training的方法一樣是用gradient descnt。f(x)=。我的w的值一開始是設為

[-1,1,-1,1,30]([前四天,前三天,前二天,前一天,常數])，iteration設為20000次，我會這樣設是因為經過不斷的測試，這樣設會有比較好的表現，而g的值會在CountLoss中被計算出來。CountLoss的作法跟basic是相同的。

另外，因為UMC的股價比較低(約60塊)，所以我在處理資料的時候，會先呈上10倍，在進行train跟predict的動作。還有因為MTK跟UMC的data在價錢上有差異，所以我用不同的learning\_rate去train這兩家公司，分別是跟。

**第二部分:**

接下來的部分就跟basic很像了，用三家公司1/1到10/14的data去做訓練，MTK跟UMC的data當作X1跟X2，TSMC的data當作Y，f(x)=。w的值一開始是設為[0,0,0]，learning\_rate都是，iteration為20000。

最後再將第一部分預測出來的資料，乘上第二部分算出來的w，就可以得到10/15~11/11 TSMC的股價。