**Summary for Lecture 3**

1. Loss Function
2. Loss Function quantifies our unhappiness with the scores across the training data
3. Multiclass SVM loss

* has this form:
* and Loss over full dataset is average
* We cannot find unique W parameter which makes loss zero by this function

1. Softmax Classifier

* Scores: unnormalized log probabilities of the classes

1. SVM vs Softmax

* When data is jiggled a bit and score’s changed, SVM loss doesn’t change but Softmax changes]

1. Regularization

* While ‘data loss’ predicts the loss on training data, ‘regularization’ works on test data for model simplicity
* We can simply add the product of regularization strength parameter and regularization function:
* L1, L2, Elastic net regularization is commonly use, and max norm regularization, dropout, batch normalization, stochastic depth will be appeared soon

1. Optimization
2. Optimization is the way of efficiently finding the parameters (like W) that minimize the loss function
3. Random Search: got 15.5% accuracy…
4. Follow the slope: Using gradient dW is silly because the loss is just a function of W
5. Gradient Descent