**NN-course**

Taking the Stanford course CS231N: <https://www.youtube.com/watch?v=vT1JzLTH4G4&list=PL3FW7Lu3i5JvHM8ljYj-zLfQRF3EO8sYv>

Class 1. Basic introduction to NN

Class 2. How NN have developed.

* Focus on vision
* How CNN have dramatically improved vision.

Class 3. The loss function - <https://www.youtube.com/watch?v=h7iBpEHGVNc&index=3&list=PL3FW7Lu3i5JvHM8ljYj-zLfQRF3EO8sYv>

* Assignment 1 -
* Loss Function - based on output, what needs to change to adapt the W matrix.
* Hinge Loss -
  + Multi-class SVM Loss @15:45 -> @18:30 - good example
  + Initialized with small weights initially
    - Initial loss should be C-1.
    - Minimum loss is 0
    - Maximum loss is technically infinite
  + Consider square hinge loss function - emphasize different types of errors. Can increase the impact of larger errors.
  + When W matrix is found, there are actually multiple W's, e.g. 2\*W. \*\* How does the classifier choose the correct W if there are multiple? Our error function doesn't push the selection one direction or another. As a result, the classifier may minimize the training data points and Overfits. Unfortunately, we want something more general rather than a complex W that matches.
    - Add Regulariztion - this attempts to make W "simpler". The regularization attempts to make the W solution more general. The regularization strength is lambda. A really important hyper-parameter. \*\* You can imagine the regularization adds a penalty for higher order polynomials.

L2 regularization - euclidean norm - W^2 L1 regularization - | W | Elastic Net - combinations of L1 + L2

All regularization attempts to reduce the complexity of a model. However, they each have different effects on the selected W.