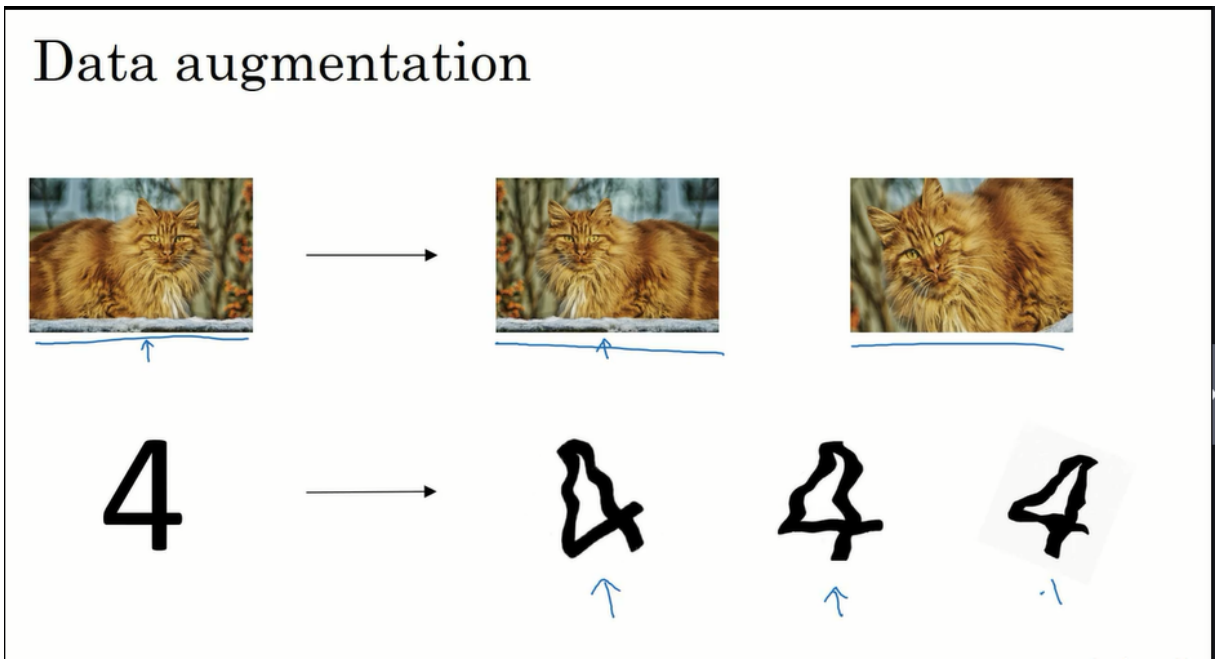
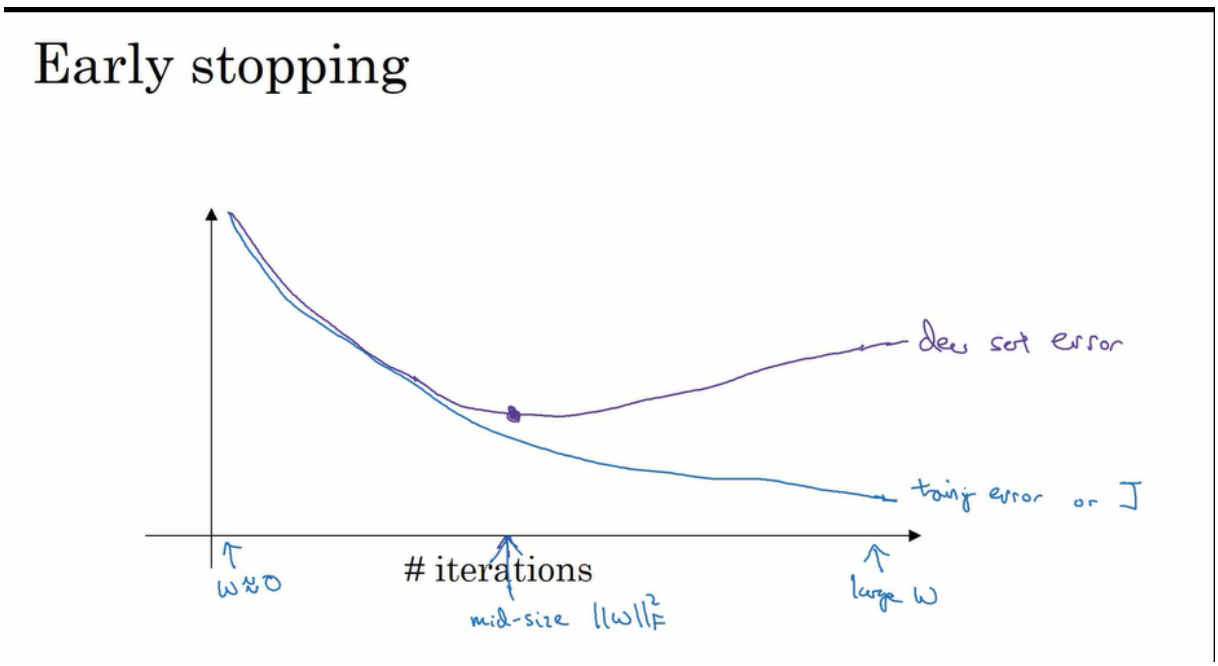
Other regularization techniques

- **Data augmentation**
  - 

- **Early stopping**
  - $at\ first\ w \approx 0$
  - $at\ last\ w\ tends\ to\ large$
  - $mid-term\ mid-dize\ \|w\|_F^2$
  - thus avoid overfitting
  - 

  - **downside**
    - Orthogonalization: fixing the 2 problems seperately

        - Optimize cost function J
        - Not overfit
    - Early stopping tends to mix the two problems:
        - minimizing the cost function J
        - in the meantime, no to overfit
    - Alternative: L2 regularization
        - computationally expensive to search for appropriate lambda
        - but iterate as long as possible
        - easier for the hyperparameter search space to decompose
- **strength**
    - get to experiment small w, mid-size w, and large w without try out lambda