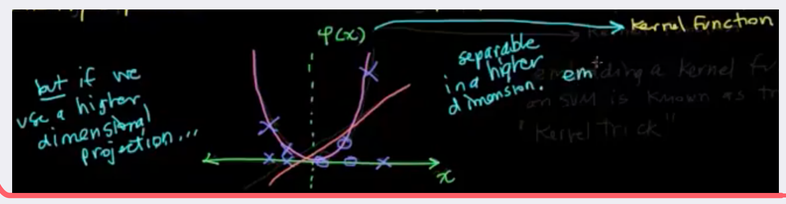
- W2
  - Introduction to Machine Learning (12A)
    - GOFAI
      - Problem solving
      - General purpose resoners
      - Operational
      - Engineered
    - ML
      - Pattern discovery
      - Specific
      - Representational
      - Data driven
      - Type
        - Supervised
          - Regression
            - Linear regresssion
            - Artificial neural networks
            - Fuzzy control
            - Regression trees
            - KNN
          - Classification
            - Logistic regression
            - SVM
            - Decision trees
            - Bayesian networks
            - KNN
        - Unsupervised
        - RL, EC
      - Features
        - Continuous, discrete
        - Numerical, Categorical
      - There is no free lunch, all have pros and cons
  - Regression (12B)
    - Pick Alpha: Too big, too small, wrong direction
    - Gradient descent
      - Very general algorithm
      - Not the usual algorithm for regression
        - Analytical, exact solution
      - illustrative
      - Good for big data
  - Tips for Regression (12C)
    - Regression
      - Consider: weighting of the feature values
      - pre-process: Mean normalization
        - $x_i' = \frac{x_i - \mu_i}{range_i}$ $range_i = x_i^{max} - x_i^{min}$ μ_i - mean of x_i
      - Categories
        - binary variables
          - Gender
          - Create different types of intercepts, but never affect the slope
          - Dummy variable, indicator variable
        - One hot encoding
  - Logistic Regression (12D)
    - Classification
    - Regression: sensitive to the outlier => Logistic sigmoid
      - $\hat{y} = \frac{1}{1+e^{-z}} = f(x)$
    - Error for logistic regression
      - $J(\theta) = -\frac{1}{n}\sum_i y_i log(\hat{y}_i) + (1-y_i) log(1-\hat{y}_i)$
    - 3 classifiers A vs B+C, B vs A+C, C vs A+B, pick the highest probability
    - Multiple binary classifiers is often superior to using a non-binary classifier
  - Support Vector Machines (12E)
    - Discriminant: separator, discriminator, between classes
    - Hinge loss function
    - SVM: Creates the wide margin
      - Support vectors, always at least 2
      - c: soft margin
      - Solved with quadratic programming(optimization)
      - Linearly separable: Separate perfectly by a hyper-plane
      - Non-linearly separable
      - 

      - Kernal trick