**L1 Supervised Learning (Jesus Please finish the basics)**

1. Linear Regression
   1. Model
   2. Cost func (OLS)
   3. Gradient Descent
      * Derivation
   4. Batch GD (LMS) (each step)
   5. Stochastic GD (each step)
   6. Matrix Derivatives
      * Gradient
      * Trace
        1. Formula
        2. Properties
        3. Combined Properties
   7. Normal Equation
      * Formula
      * Derivation
   8. Probabilistic Interpretation
      * Model
        1. + Gaussian
      * Probabilistic Model
      * Likelihood Func
      * MLE
   9. Locally Weighted Linear Regression
      * Original Linear Regression
        1. S1:
        2. S2:
      * Weighted Linear Regression
        1. S1:
           1. bandwidth param

Small weight

Large weight

* + - 1. S2:

1. Classification & Logistic Regression
   1. LogReg
      * Formula
      * Sigmoid func
        1. Derivative
           1. Derivation
      * MLE
        1. Assumptions
        2. Prob Model
        3. Likelihood Func
        4. Log likelihood
        5. MLE(SGD)
        6. SGA
   2. GD - Newton’s Method
      * Newton-Raphson Method
      * Hessian
      * Newton VS GD
        1. Adv: faster convergence, fewer iterations
        2. Disadv: one iteration is expensive as hell (we need to find & invert a Hessian)
2. Generalized Linear Models (GLM)
   1. Summary
      * Regression:
      * Classification:
      * Regression + Classification Generalized Linear Models (GLMs)
   2. Exponential family
      * General form of distribution
      * Notations
        1. natural parameter (i.e. canonical parameter)
        2. sufficient statistic (usually )
        3. log partition function
        4. normalization constant (ensure that )
        5. : fixed choice that defines a family/set of distributions that is parametrized by
           1. As we vary , we then get different distributions within this family
      * e.g. Bernoulli Distribution
      * e.g. Gaussian Distribution (set )
      * Other members of Exponential Family:
        1. Multinomial
        2. Poisson: modeling count-data
        3. Gamma: modeling continuous, non-negative random vars (e.g. time intervals)
        4. Beta & Dirichlet: distributions over probabilities
3. Constructing GLMs
   1. 3 assumptions
      * ( linearly)
   2. Derivation examples
      * e.g. OLS
      * e.g. LogReg
      * e.g. Softmax Regression
        1. Problem setting
        2. Define parameters:
        3. Linear Transformation (encoding):
        4. Indicator Function:
        5. Derivation of Exponential Family
        6. Derivation of Distribution
           1. Derive
           2. Derive distribution
        7. Hypothesis
        8. Log likelihood

**L2 Generative Learning Algorithms**

1. Bayes Summary
2. Gaussian Discriminant Analysis
3. Naïve Bayes