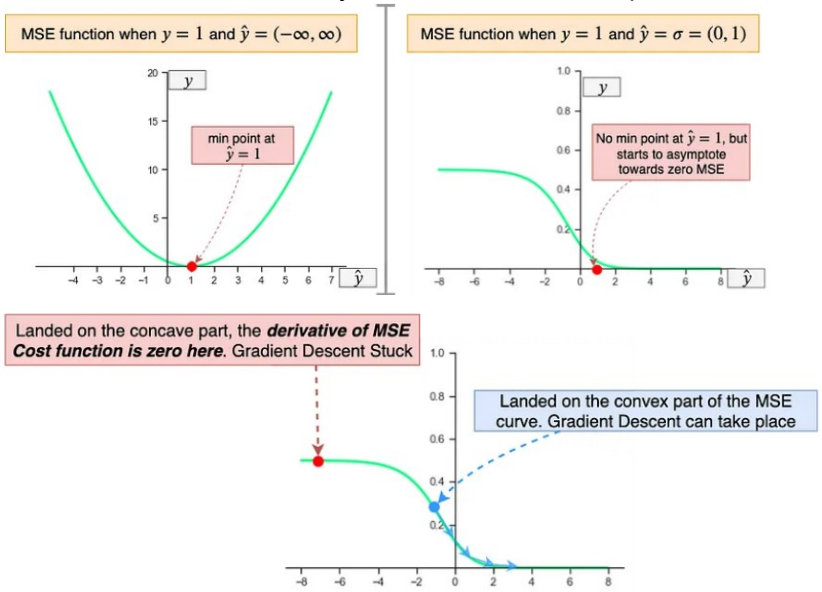
Logistic Regression

1. Code for Logistic Regression
2. What is the Hypothesis
    a. $h_\theta(x) = g(\theta^T X) = 1/ 1+ e^{(-\theta^T X)}$
3. Loss function:
    a. J =
4. Update rule
    a. $\theta = \theta + lr * \Sigma (y^i - h_\theta(x^i))x^i_j$
    b. same as Linear regression. **Why**? - because both belong to GLM (Generalized Linear Models) and for all GLMs this is the update rule.
5. There could be many function that ranges from 0 to 1 then, why Logistic is chosen?
    a. Comes from GLM – comes from exponential family of Bernoulli
6. Why is MSE not a good loss for Logistic Regression?
    a. Because loss is not convex (??), outliers sensitive (??), that's why log(loss) (??)
7. Link between Logistic Regression and Bernoulli distribution
    a. Check Exponential family and Generalized Linear Model
8. How does the loss curve for Cross entropy look?
9. What does the "minus" in cross-entropy mean?
10. Can you use MSE for evaluating your classification problem instead of Cross entropy
    a. Not recommended
    b. MSE loss becomes too low while CE loss has log term and sice values are in range of 0 to1, log loss is higher -> better updates to gradient descent with CE loss as compared to MSE loss
    c. Values can shoot 1 or below 0
    d. MSE is non-convex for binary classification -> not optimum (sensitive to initialization)



    e. MSE assumes normal distribution but for Logistic regression data comes from Bernoulli Distribution.

    f. Decision boundary Becomes sensitive to outliers.
11. Proof: derive gradient descent for Logistic regression
12. Impact of outliers
    a. As compared to Linear Regression, Logistic regression is less sensitive to outliers
13. Difference between Linear Regression and Logistic regression
    a. Linear regression: deps and indeps have linear relationship || Logistic regression – linear relationship between deps and log-odds of indep
    b. Linear regression – predicts the value of dep || Logistic regression – predicts the probability of dep event occulting
    c. Linear Regression: assumes residuals are normally distributed || Logistic Regression – binomial error
    d. Linear Regression – MSE as loss || Logistic Regression: CE as loss
    e. Error minimization method: Linear Regression (minimize OLS) || Logistic Regression (maximize Likelihood)