Assumptions in ML

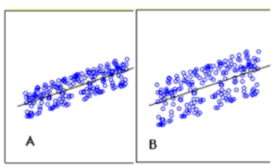
1. Role of IID – independent & identically distributed
    a. Meaning errors ξ_i are normally distributed. $\xi i \sim N(0, \sigma 2)$ – Important role in deriving MSE as natural choice of cost function for Linear Regression.
    b. When we divide dataset into train-val-test, we have this underlying assumption of IID. If they don't have IID => Poor Performance

Linear Regression

1. Code for Linear Regression
2. Assumptions:
    a. Indeps and deps are linear
    b. Residual/error: Normally distributed with zero mean and a constant variance
    c. Indeps are independent (not correlated) of each other
    d. Residuals are independent of each other
    e. Errors are iid
3. What is the Hypothesis
    a. $h_\theta(x) = \theta^T X = \Sigma \theta_i X_i$
4. Loss function:
    a. $J = (Y-X\theta)^T(Y-X\theta)$
5. Update rule
    a. $\theta = \theta + lr*\Sigma (y^i - h_\theta(x^i))x^i_j$
    b. same as logistic regression. **s**? - because both belong to GLM (Generalized Linear Models) and for all GLMs this is the update rule.
6. Proof - Why is MSE a good loss for Linear Regression?
    a. It gets derived as a natural choice by exploiting IID assumption.
    b. On maximizing Log Likelihood (differentiate and set to zero), meant reducing MSE
7. What does it mean a function is convex?
    a. F''(x) (aka Hessian) > 0
    b. domain is real
8. Proof- MSE for Linear Regression is convex
9. What are loss function in Linear Regression and Regularized Linear Regression
    a. Linear Regression: $J = (Y-X\theta)^T(Y-X\theta)$
    b. Regularized Linear Regression: $J = (Y-X\theta)^T(Y-X\theta) + \lambda\theta^T\theta$
10. Proof: Derive $\theta = (X^TX)^{-1}X^TY$ for Linear Regression
11. What if, $(X^TX)^{-1}$ is not invertible?
    a. Means there are correlated features - multicollinearity.
    b. So possible solution is: use ridge regression (L2 regularization), LASSO regression, SVD
    c. Pseudo inverse using SVD
12. Proof: Derive $\theta = (X^TX+ \lambda I)^{-1}X^TY$ for Regularized Linear Regression
13. Link between Linear Regression and Gaussian distribution
    a. Check Exponential family and Generalized Linear Model

14. Impact of outliers on Linear Regression
    a. Highly sensitive to outliers
15. Correlation coefficient formula
    a. Ranges from -1 to 1
    b. $$R^2 = 1 - \frac{\text{SS Residual}}{\text{SS Total}} \quad \text{TSS} := \sum_{i=1}^{N}(y_i - \bar{y})^2 . \text{RSS} := \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$
16. If time permits, read about Locally Weighted Linear Regression. Can we use GD in this case? Why or why not? Also implement it via code.
    a. Closed form solution: $(X^T wX)^{-1} X^T wY$
17. Difference between Linear Regression and locally weighted linear regression
    a. Linear regression – parametric while locally weighted linear regression – non-parametric
18. Code implementation of locally weighted linear regression.

2. X-axis is the independent variable, and Y-axis is the dependent variable.



A) A has a higher sum of residuals than B

B) A has a lower sum of residual than B

C) Both have the same sum of residuals

D) None of these

Solution: (C)

19. The sum of residuals will always be zero; therefore, both have the same sum of residuals.
20. Why is it important that the error terms have equal variance in linear regression?
    a. the assumption that the error terms (or residuals) have equal variance is known as **homoscedasticity**. **Ordinary Least Squares (OLS)** estimators are unbiased regardless of whether homoscedasticity holds. However, if the error terms have equal variance (homoscedasticity), the OLS estimators are also the Best Linear Unbiased Estimators (BLUE). This means that among all linear and unbiased estimators, OLS has the smallest variance, making it the most efficient.
21. Drawback of R-squared?
    a. Its value keep on increasing if we add predictors (features) -> even though actually it might be leading to overfitting (in those case use **Adjusted R-squared**)
22. Normalization vs Standardization
    a. Normalization: data in range 0 to 1
    b. Standardization: mean=0, std=1
23. R-squared is a measure that represents the proportion of variance explained by the model. Explain