

Week 1

Summary Questions

- Define a hyperplane in a way that helps you understand it conceptually.
 - A divider for a high-dimensional space that separates two classes of data.

Week 2

Synthesis Questions

Linear Regression

- What is a feature in this context?
 - A feature in linear regression is a variable that is used as input to form a linear model in order to predict the output.
- What are the significant of the β terms within the modified $y = mx + b$ equations described in the article?
 - They are the weights for each feature of the linear model and hence is important to the model's accuracy.
- What is SSE:
 - **How is it calculated?**
 - It is the sum of the squared errors of the model where error is the difference between the predicted value and the actual value.
 - **What can it tell you about the values you chose for β**
 - They can tell you how each feature of the input data affects the output, i.e. what features are more significant to the model.
 - **If you modify the β_1 term and the SSE goes up, was that a good modification?**
 - No, because SSE is the sum of the squared errors, hence if SSE goes up, the error went up and the model is less accurate.
- Write out the linear regression formula (involving β) when you wish to estimate the impact of age, height, and weight of someone regarding their marital status.
 - Marital Status = $\beta_0 + \beta_1 * \text{Age} + \beta_2 * \text{Height} + \beta_3 * \text{Weight}$
- What are the 4 assumptions described that you need to confirm before using linear regression?
 1. Linear relationship between the input and output
 2. Constant variance of the data - homoscedasticity
 3. Independence of errors
 4. Lack of perfect multicollinearity - features should be linearly independent of each other
- What is homoscedasticity? What is heteroscedasticity?
 - Homoscedasticity is when the variance of the data is constant, while heteroscedasticity is when the variance of the data is not constant.

For linear regression, the data needs to be homoscedastic.

SVM

- What are some use cases for an SVM? What does it do?
 - SVM is useful for classification problems by finding a hyperplane that classifies different classes of data.
- What is the margin?
 - Margin is the distance between the hyperplane and the closest data points.
- Why does the margin need to be maximized? What does this allow for?
 - Because then the hyperplane is as far as possible from the closest data points, hence allowing more data points in between the classification hyperplane and the data points.
- What are the support vectors?
 - They are the data points that are closest to the hyperplane and are used to define the margin.
- What is the difference between a hard and soft margin SVM?
 - Hard margin SVM is when the data points are linearly separable, while soft margin SVM is when the data points are not linearly separable. It means that, in a hard margin, all data points are on the correct side of the hyperplane, while in a soft margin, some data points are on the wrong side of the hyperplane.

SVM Math

- $w \cdot x - b = -1$ defines a margin of support vectors.
- What are the w variables in the equation?
 - The weight values applied to each feature of x
- What is the equation for the decision boundary?
 - $\vec{w} \cdot \vec{x} - b = 0$
- What is the size of the margin in terms of the vector w

The entire margin across the decision boundary: $\frac{2}{\|\vec{w}\|}$

- What needs to be minimized, and what are the constraints for finding the optimal hyperplane decision boundary?
 - Minimize $\|\vec{w}\|$ in order to maximize the margin
 - The constraints are:

$$y_i(\vec{w}_i \cdot \vec{x}_i - b) \geq 1 : \begin{cases} y_1 = 1 \implies \vec{w} \cdot \vec{x} - b \geq 1 \\ y_1 = -1 \implies \vec{w} \cdot \vec{x} - b \leq -1 \end{cases}, \quad \forall i \in [1, \dots, N]$$

PCA

- What does PCA do?

- It reduced the dimensions of a dataset by trends among different features and combining them into principal components
 - It is useful when trying to visualize data beyond 3D
- What is a principal component?
 - A eigenvector of covariance matrix that is importance-ranked by their eigenvalues
 - They are combined forms of original dimensions in the dataset to reduce dimensions