

Homework 5

Cauchy-Schwarz Inequality

The Cauchy-Schwarz inequality is a mathematical inequality that establishes an upper limit for the absolute value of the scalar product between two vectors in terms of their norms.

Formal Statement

Given the inequality:

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

where:

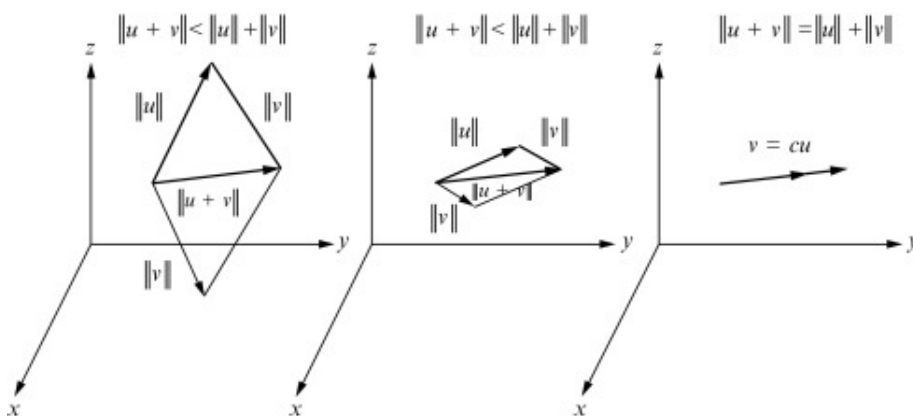
- $\langle \mathbf{u}, \mathbf{v} \rangle$: scalar product between two vectors \mathbf{u} and \mathbf{v}
- $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$: norms (or lengths) of the vectors \mathbf{u} and \mathbf{v} .

This inequality states that the scalar product between two vectors cannot exceed the product of their norms. In other words, it measures how "aligned" the two vectors are: the scalar product is maximal if the vectors point in the same (or opposite) direction and decreases as the vectors form a wider angle.

Geometric Interpretation

Imagine two vectors \mathbf{u} and \mathbf{v} in Euclidean space:

- If they are parallel (same direction), the scalar product $\langle \mathbf{u}, \mathbf{v} \rangle$ is exactly equal to $\|\mathbf{u}\| \|\mathbf{v}\|$.
- If they form an angle, the scalar product will be less than $\|\mathbf{u}\| \|\mathbf{v}\|$.
- If they are orthogonal, the scalar product is zero, so $|\langle \mathbf{u}, \mathbf{v} \rangle| = 0 \leq \|\mathbf{u}\| \|\mathbf{v}\|$.



Demonstration

Consider the quadratic function with respect to a parameter t defined as:

$$f(t) = \|\mathbf{u} + t\mathbf{v}\|^2$$

Expanding $f(t)$, we obtain:

$$f(t) = \langle \mathbf{u} + t\mathbf{v}, \mathbf{u} + t\mathbf{v} \rangle = \|\mathbf{u}\|^2 + 2t\langle \mathbf{u}, \mathbf{v} \rangle + t^2\|\mathbf{v}\|^2$$

This function is a parabola with respect to t , and we can write it in the standard form:

$$f(t) = at^2 + bt + c$$

where:

- $a = \|\mathbf{v}\|^2$
- $b = 2\langle \mathbf{u}, \mathbf{v} \rangle$
- $c = \|\mathbf{u}\|^2$

Since $f(t) \geq 0$ for every t , the discriminant of this quadratic function must be less than or equal to zero to ensure that the parabola does not intersect the t -axis.

Now we can calculate the discriminant by substituting the values of a , b , and c :

- $b = 2\langle \mathbf{u}, \mathbf{v} \rangle \Rightarrow b^2 = (2\langle \mathbf{u}, \mathbf{v} \rangle)^2 = 4\langle \mathbf{u}, \mathbf{v} \rangle^2$
- $a = \|\mathbf{v}\|^2$ and $c = \|\mathbf{u}\|^2 \Rightarrow 4ac = 4\|\mathbf{u}\|^2\|\mathbf{v}\|^2$

Thus, the discriminant becomes:

$$\Delta = 4\langle \mathbf{u}, \mathbf{v} \rangle^2 - 4\|\mathbf{u}\|^2\|\mathbf{v}\|^2$$

To satisfy $f(t) \geq 0$ for all t , we need $\Delta \leq 0$, so:

$$4\langle \mathbf{u}, \mathbf{v} \rangle^2 - 4\|\mathbf{u}\|^2\|\mathbf{v}\|^2 \leq 0$$

Dividing by 4, we obtain:

$$\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2$$

We can conclude by writing:

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

Independence of Random Variables

Two random variables X and Y are considered independent if the occurrence or value of one variable does not provide any information about the occurrence or value of the other. In other words, knowing the outcome of Y does not change the probability distribution of X , and vice versa.

Formal Definitions of Independence

1. Independence in Terms of Conditional Distributions

Independence implies that the conditional distribution of X given Y remains the same as the marginal distribution of X . This means that the value of Y has no influence on the distribution of X , preserving its probability distribution as if Y were unknown:

$$f_{X|Y=y_j} = f_X(x_i)$$

This condition tells us that knowing $Y = y_j$ does not affect the likelihood of any specific outcome $X = x_i$.

2. Independence in Terms of Probabilities

For independent random variables, the conditional probability of an event involving X , given that an event involving Y has occurred, is the same as the marginal probability of the event involving X .

Formally:

$$P(X = x_i | Y = y_j) = P(X = x_i)$$

This relation indicates that the probability of $X = x_i$ does not change based on knowledge of $Y = y_j$.

3. Independence in Joint Distributions

In the context of joint distributions, independence between X and Y is characterized by the product of their individual (marginal) distributions. This holds for both discrete and continuous random variables and can be expressed as:

$$f_{XY}(x_i, y_j) = f_X(x_i) \cdot f_Y(y_j)$$

This equation signifies that the joint probability (or joint density) of X and Y occurring together is simply the product of their marginal probabilities (or densities), confirming that there is no dependency between them.

Correlation

Correlation is a statistical measure that describes the strength and direction of a linear relationship between two random variables. It quantifies how much two variables tend to move together in a straight-line pattern. The correlation coefficient, typically denoted by r , can take values in range $\{-1, 1\}$:

- $r = 1$ indicates a **perfect positive correlation**: as one variable increases, the other increases proportionally in a linear manner.
- $r = -1$ indicates a **perfect negative correlation**: as one variable increases, the other decreases in a perfectly linear fashion.
- $r = 0$ indicates **no linear correlation**: there is no apparent linear relationship between the two variables.

Formula for Calculating Correlation

The correlation coefficient r is calculated as follows:

$$r = \frac{\sum_{i,j} (x_i - \bar{x})(y_j - \bar{y})}{\sqrt{\sum_j (x_j - \bar{x})^2 \sum_j (y_j - \bar{y})^2}} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y}$$

where:

- $\sigma_{x,y}$ represents the **covariance** between X and Y , which measures the extent to which the two variables vary together.
- σ_x and σ_y are the **standard deviations** of X and Y , respectively, indicating how much each variable deviates from its mean.

The formula divides the covariance between X and Y by the product of their standard deviations, normalizing the measure to fall within the range of -1 to 1 .

Relationship Between Independence and Correlation

In statistics, there is an important relationship between independence and correlation:

1. Independence Implies Zero Correlation

If two variables X and Y are **independent**, then they have a correlation of zero. This means that if there is no influence or dependency between the two variables, they do not have a linear relationship either. Independence ensures that knowing the value of X provides no information about Y , and vice versa, which results in no linear correlation.

2. Zero Correlation Does Not Imply Independence

However, the viceversa is not true: a correlation of zero ($r = 0$) does not imply that two variables are independent. A correlation of zero only indicates that there is no linear relationship between X and Y . The variables could still be related in a non-linear way. For example, X and Y may have a **quadratic relationship** (such as $Y = X^2$), which is not captured by the correlation coefficient. Thus, even though $r = 0$, X and Y might still be dependent.

Regression Coefficients

In linear regression, we aim to model the relationship between an independent variable X and a dependent variable Y using a linear equation:

$$Y = a + bX$$

The **least squares method** is used to find the values of a and b that minimize the sum of squared differences (errors) between the observed values Y and the predicted values $\hat{Y} = a + bX$.

Deriving the Slope (b) and Intercept (a)

1. Slope (b)

The slope b in linear regression is derived based on minimizing the **sum of squared errors** between observed values and predicted values along a line:

- We want to minimize the **sum of the squared vertical distances** between each observed Y_i and its corresponding predicted $\hat{Y}_i = a + bX_i$. The **error** for each point i is $e_i = Y_i - (a + bX_i)$. To minimize the overall error, we minimize the sum of squared errors $\sum e_i^2$:

$$\text{Error} = (Y_i - (a + bX_i))^2$$

- To find the values of a and b that minimize the error, we take the partial derivatives of the error function with respect to a and b , and set them to zero.

For the slope b :

1. **Partial Derivative with respect to b :** Setting this derivative to zero, we obtain:

$$\sum (Y_i - (a + bX_i))X_i = 0$$

2. **Solving for b :** Replacing a with $\bar{Y} - b\bar{X}$, and simplifying, we get the formula for b in terms of deviations from the mean:

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

3. We can notice that:

- $\text{Cov}(X, Y) = \frac{1}{n} \sum (X_i - \bar{X})(Y_i - \bar{Y})$
- $\text{Var}(X) = \frac{1}{n} \sum (X_i - \bar{X})^2$

We can conclude by writing:

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

This formula calculates the slope by finding the ratio of the covariance of X and Y to the variance of X .

2. Intercept (a)

Once we have the slope b , the intercept a can be calculated using the means of X and Y :

$$a = \bar{Y} - b\bar{X}$$

This formula ensures that the regression line passes through the mean point (\bar{X}, \bar{Y}) , balancing the line based on the average values of X and Y .

Relationship with R^2 (Coefficient of Determination)

The coefficient of determination, R^2 , measures the proportion of the variance in the dependent variable Y that is predictable from the independent variable X . It provides an indication of the goodness of fit of the regression model. R^2 is calculated as:

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{tot}}} = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

where:

- SS_{reg} is the **regression sum of squares**, representing the variance explained by the model.
- SS_{tot} is the **total sum of squares**, representing the total variance in Y .
- SS_{res} is the **residual sum of squares**, representing the unexplained variance in Y after accounting for X .

The R^2 value ranges from 0 to 1:

- An $R^2 = 1$ indicates a perfect fit, where all points lie exactly on the regression line.

- An $R^2 = 0$ indicates that the regression model does not explain any of the variance in Y .

The relationship between R^2 and the regression coefficients a and b demonstrates how well the regression line fits the observed data. A higher R^2 value suggests that the line, defined by the coefficients a and b , accurately represents the linear relationship between X and Y .