Artificial Intelligence (Machine Learning & Deep Learning) [Course]

Week 9 – FAST BACKWORD LOOK – Various Topics

[See examples / code in GitHub code repository]

It is not about Theory, it is 20% Theory and 80% Practical – Technical/Development/Programming [Mostly Python based]

What is Covariance?

It’s a statistical term demonstrating a systematic association between two random variables, where the change in the other mirrors the change in one variable.

Definition and Calculation of Covariance

Covariance implies whether the two variables are directly or inversely proportional.

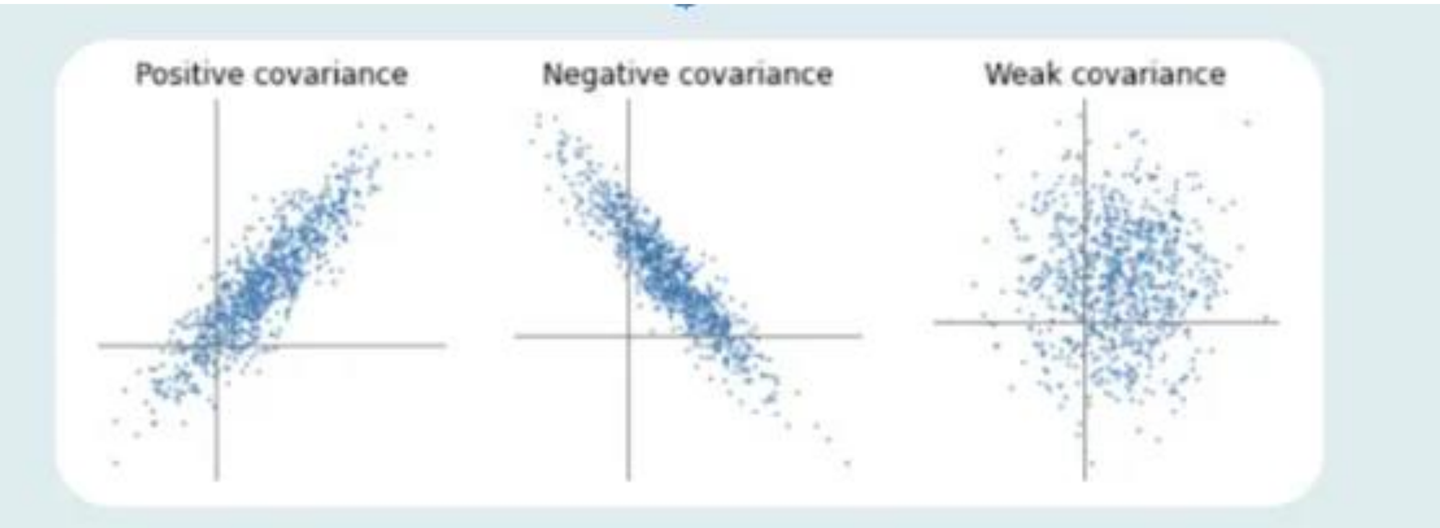
The covariance formula determines data points in a dataset from their average value. For instance, you can compute the Covariance between two random variables, X and Y, using the following formula:

$$\sigma(x, y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

Reference:
https://www.analyticsvidhya.com/blog/2023/07/covariance-vs-correlation/

Positive, Negative, and Zero Covariance

The higher the number, the more reliant the relationship between the variables. Let's comprehend each variance type individually:



Positive Covariance

- If the relationship between the two variables is a positive covariance, they are progressing in the same direction. It represents a direct relationship between the variables. Hence, the variables will behave similarly.
- The relationship between the variables will be positive Covariance only if the values of one variable (smaller or more significant) are equal to the importance of another variable.

Negative Covariance

- A negative number represents negative Covariance between two random variables. It implies that the variables will share an inverse relationship. In negative Covariance, the variables move in the opposite direction.
- In contrast to the positive Covariance, the greater of one variable correspond to the smaller value of another variable and vice versa.

Continued..

Zero Covariance

Zero Covariance indicates no relationship between two variables.

Significance of Covariance in Assessing Linear Relationship

- Covariance is significant in determining the linear relationship between variables. It suggests the direction (negative or positive) and magnitude of the relationship between variables.
- A higher covariance value indicates a strong linear relationship between the variables, while a zero covariance suggests no ties.

Limitations and Considerations of Covariance

- The scales of measurements influence the Covariance and are highly affected by outliers. Covariance is restricted to measuring only the linear relationships and doesn't apprehend the direction or strength.
- Moreover, comparing covariances across various datasets demand caution due to different variable ranges.

Reference:
https://www.analyticsvidhya.com/blog/2023/07/covariance-vs-correlation/

What is Correlation?

Unlike Covariance, correlation tells us the direction and strength of the relationship between multiple variables. Correlation assesses the extent to which two or more random variables progress in sequence.

Definition and Calculation of Correlation Coefficient

Correlation is a statistical concept determining the relationship potency of two numerical variables. While deducing the relation between variables, we conclude the change in one variable that impacts a difference in another.

When an analogous movement of another variable reciprocates the progression of one variable in some manner or another throughout the study of two variables, the variables are correlated.

The formula for calculating the correlation coefficient is as follows:

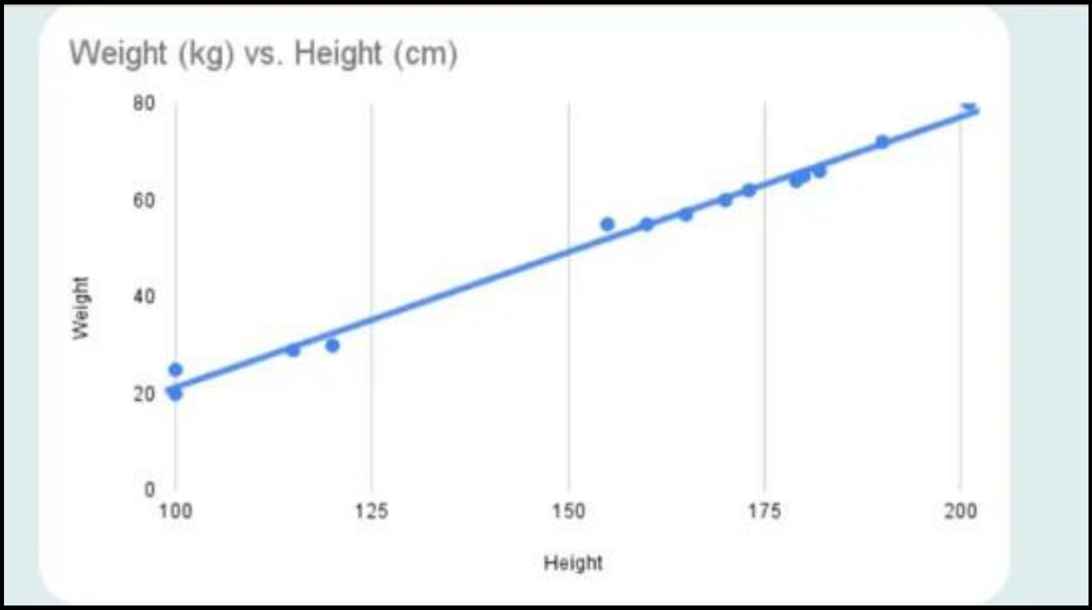
$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Reference:
https://www.geeksforgeeks.org/artificial-intelligence/llms-vs-generative-ai-dissecting-the-spectrum-of-ai-powered-creativity/

Positive, Negative, and Zero Correlation

If the variables are directly proportional to one another, the two variables are said to hold a positive correlation. This implies that if one variable's value rises, the other's value will exceed. An ideal positive correlation possesses a value of 1.

Here's what a positive correlation looks like:



In a negative correlation, one variable's value increases while the second one's value decreases. A perfect negative correlation has a value of -1.

The negative correlation appears as follows:

Continued..

Just like in the case of Covariance, a zero correlation means no relation between the variables. Therefore, whether one variable increases or decreases won't affect the other variable.

Strength and Direction of Correlation

Correlation assesses the direction and strength of a linear relationship between multiple variables. The correlation coefficient varies from -1 to 1, with values near -1 or 1 implying a high association (negative or positive, respectively) and values near 0 suggesting a weak or no correlation.

To summarize the differences, here's a table you must glance through:

Difference Grounds	Covariance	Correlation
Meaning	Covariance means two variables directly or inversely depend on one another.	Two variables are said to be in correlation if the change in one affects the other variable.
Values	Lie between -infinity to +infinity	Values lie between -1 to 1
Unit	It's a product of the unit of variables	It's a unit-free measure
Change in Scale	Even minor changes in scale affect Covariance	There won't be any change in correlation because of the scale
Measure of	Correlation	The scaled version of Covariance
Application	Market Research, Portfolio Analysis, and Risk Assistance	Medical Research, Data Analysis, and Forecasting

Reference:
https://www.analyticsvidhya.com/blog/2023/07/covariance-vs-correlation/

Outlier Detection, and Data Interpretation

What are Outliers?

We all have heard of the idiom 'odd one out which means something unusual in comparison to the others in a group. Similarly, an Outlier is an observation in a given dataset that lies far from the rest of the observations. That means an outlier treatment is vastly larger or smaller than the remaining values in the set.

What Do They Affect?

In statistics, we have three measures of central tendency namely Mean, Median, and Mode. They help us describe the data.

- ❑ Mean is the accurate measure to describe the data when we do not have any outliers present.
- ❑ Median is used if there is an outlier in the dataset.
- ❑ Mode is used if there is an outlier AND about ½ or more of the data is the same.

Mean' is the only measure of central tendency that is affected by the outlier treatment which in turn impacts Standard deviation.

Reference:
https://www.analyticsvidhya.com/blog/2023/07/covariance-vs-correlation/

Continued...

Detecting Outliers

If our dataset is small, we can detect the outlier by just looking at the dataset. But what if we have a huge dataset, how do we identify the outliers then? We need to use visualization and mathematical techniques.

Below are some of the techniques of detecting outliers

- Boxplots [Sample Code]
- Z-score
- Inter Quantile Range(IQR)

To summarize the differences, here's a table you must glance through:

Reference:
https://www.analyticsvidhya.com/blog/2023/07/covariance-vs-correlation/

How to Handle Outliers?

Step 1: Trimming/Remove the outliers

In this technique, we remove the outliers from the dataset. Although it is not a good practice to follow.

Step 2: Quantile Based Flooring and Capping

In this technique, the outlier is capped at a certain value above the 90th percentile value or floored at a factor below the 10th percentile value.

Step 3: Mean/Median Imputation

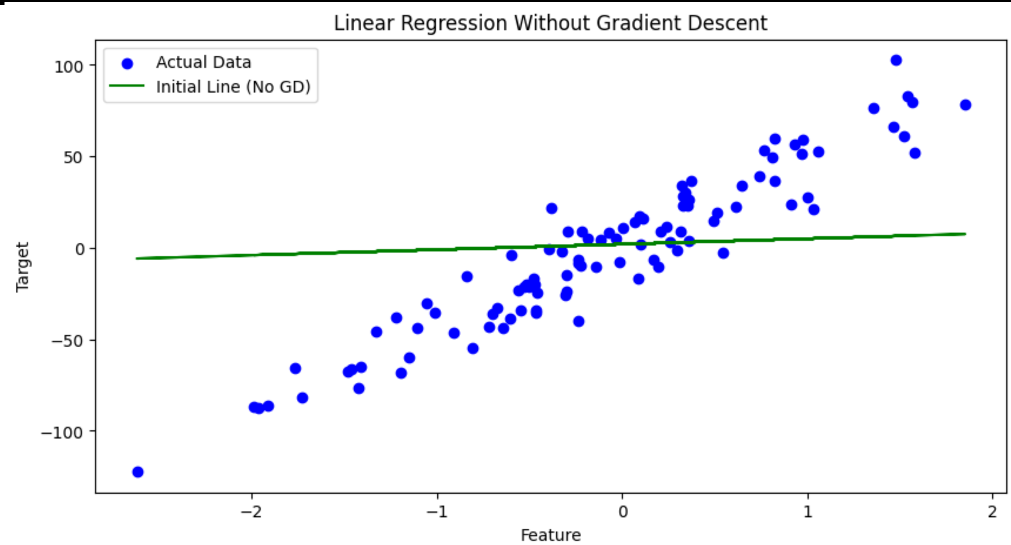
As the mean value is highly influenced by the outlier treatment, it is advised to replace the outliers with the median value.

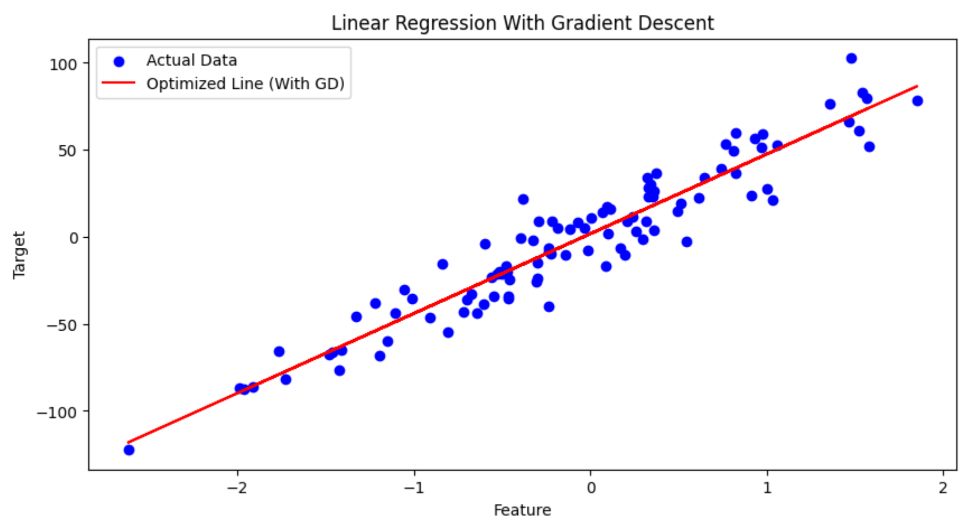
Reference:

https://www.analyticsvidhya.com/blog/2021/05/detecting-and-treating-outliers-treating-the-odd-one-out/

Gradient Descent in Linear Regression

Gradient descent is a optimization algorithm used in linear regression to find the best fit line tohe data. It works by gradually by adjusting the line's slope and intercept to reduce the difference between actual and predicted values. This process helps the model make accurate predictions by minimizing errors step by step.

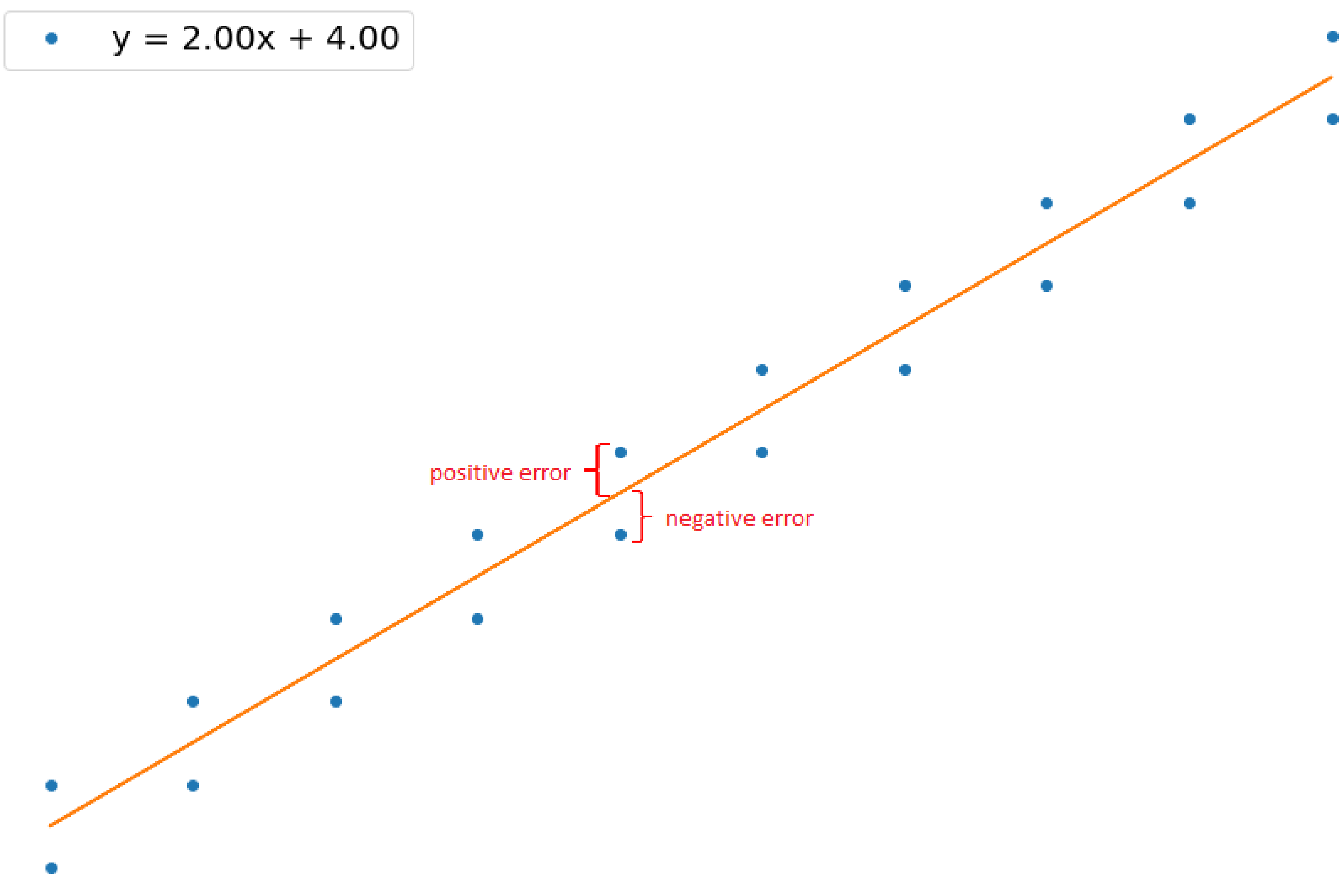




Reference:
https://www.geeksforgeeks.org/machine-learning/gradient-descent-in-linear-regression

Regularization (L1, L2)

One way to reduce the complexity of the model, is to push the coefficients towards zero. A further benefit of pushing the coefficients towards zero (if the coefficients for some features actually achieve zero) is to practically remove the impact of features that have no predictive power or are multicollinear with other features. A technique called regularisation is used to achieve this.



Reference:
https://towardsdatascience.com/simple-regularized-linear-and-polynomial-regression-37d0d634ece3

Continued...

Ridge Regression — Also called the L2 norm. This is called the L2 Norm because the ϴ values are squared in the penalty.
Lasso Regression — Also called the L1 norm. This is called the L1 Norm because the ϴ values are absolute values (Degree of 1) in the penalty.
ElasticNet Regression — Where a combination of the Ridge and Lasso regularisation techniques are used.
There can be many other techniques (e.g., taking 0.5, 3 or 4 as the degree of exponent for ϴ) which are denoted by L*p*.
However, we shall only delve into the 3 most common techniques.

Ridge Regression (L2 Norm)

The penalty to the cost function is added using the following equation:
RSS = Σ(Ÿ — Y)2
Since **Ÿ = w0 + w1x1…wnxn**
RSS = Σ((w0+w1x1…wnxn) — Y)2
Regularisation just adds a penalty to this equation.
Ridge Regression = (Σ ((w0 + w1x1…wnxn) — Y)2 + λ Σ(w1…wn)2
Where λ is a factor that we manually set to multiply with the sum of the coefficients. With this added penalty, our optimizer function, will continue to look for weights to minimize the slope i.e., the higher the coefficient, the more penalty will be added and the closer the coefficient gets to zero. In features not showing any collinearity with the Y variable, the coefficients will be close to zero, preventing our model from picking up random noise and causing overfitting.

Lasso Regression (L1 Norm)

The penalty to the cost function is added using the following equation:
Lasso Regression = (Σ ((w0 + w1x1…wnxn) — Y)2 + λ Σ|w1…wn|
Similar to the ridge regression, the added penalty, pushes the coefficients to zero. However, since the penalty is not magnified by squaring — there is the possibility that the coefficients for certain features actually achieve zero. Features showing no correlation at all with the Y variable will achieve 0. Also, if there are a bunch of multicollinear features and one of them has already been assigned weights accounting for the predictive power of the bunch, then the rest of the features will also end up having zero coefficients and practically dropped from the model.

Transfer Learning (ResNet, VGG, Inception)

Transfer learning leverages pre-trained models like ResNet, VGG, and Inception to accelerate and improve performance on new, related tasks. These models are trained on massive datasets (like ImageNet) and their learned features can be adapted to different image recognition tasks.

How Transfer Learning Works:

1. Pre-trained Models:

VGG, ResNet, and Inception are popular CNN architectures that have been trained on vast image datasets. They learn generalizable features like edges, textures, and shapes in the initial layers.

2. Adaptation:

Instead of training a new model from scratch, you can "transfer" the knowledge from these pre-trained models to your specific task.

3. Fine-tuning:

The pre-trained model's layers (especially the initial ones) are often frozen (their weights are not updated during training), and new layers are added to adapt to the new dataset and task. Fine-tuning involves adjusting the weights of the frozen layers to better suit the new task.

Key Considerations:

Dataset Similarity:

Transfer learning is most effective when the source task (the task the model was originally trained on) and the target task (your new task) are related.

Computational Resources:

Transfer learning can significantly reduce training time and resource requirements compared to training from scratch.

Model Selection:

The choice of pre-trained model depends on the complexity of your task and the characteristics of your data.

Advantages of Transfer Learning:

Faster Training: Reduced training time and faster convergence.

Improved Accuracy: Pre-trained models can offer higher accuracy, especially when the target dataset is small.

Reduced Data Requirements: Can be effective even with limited data.

Feature Extraction: Pre-trained models act as feature extractors, allowing you to focus on building the final classification/regression layers.

In essence, transfer learning allows you to "borrow" the knowledge of a pre-trained model, saving time and resources while potentially achieving better results on your specific image recognition problem.

Thank you - for listening and participating

- ❑**Questions / Queries**
- ❑**Suggestions/Recommendation**
- ❑**Ideas…..?**

Shahzad Sarwar
Cognitive Convergence
https://cognitiveconvergence.com
shahzad@cognitiveconvergence.com
voice: +1 4242530744 (USA) +92-3004762901 (Pak)