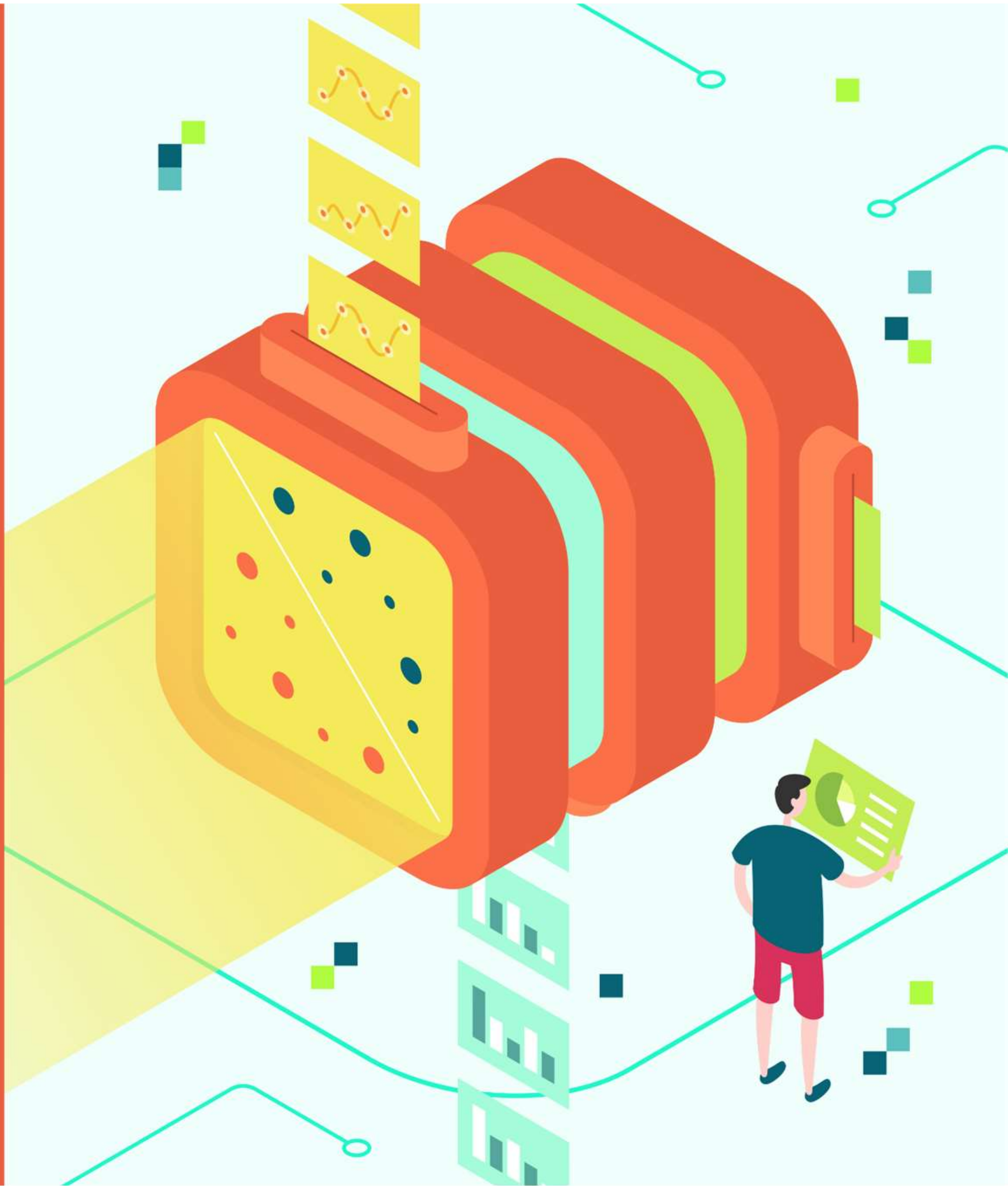
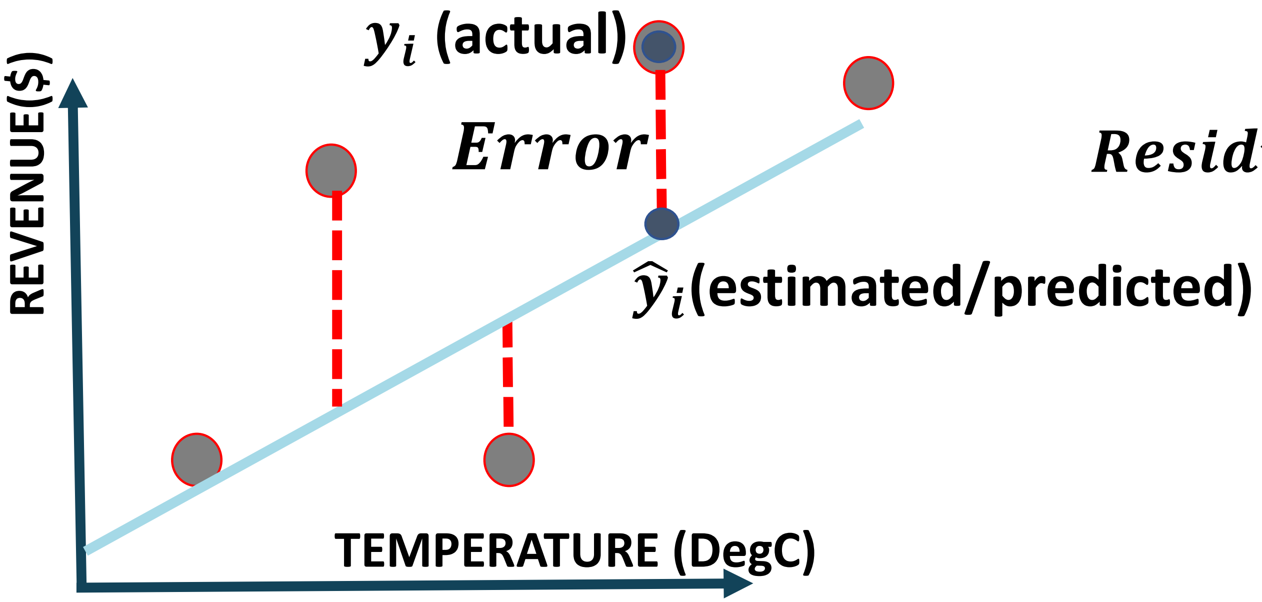
MACHINE LEARNING REGRESSION

REGRESSION METRICS



REGRESSION METRICS: HOW TO ASSESS MODEL PERFORMANCE?

- After model fitting, we would like to assess the performance of the model by comparing model predictions to actual (True) data



$$Residuals\ (Error) = \hat{y}_i - y_i$$

REGRESSION METRICS: MEAN ABSOLUTE ERROR (MAE)

- Mean Absolute Error (MAE) is obtained by calculating the absolute difference between the model predictions and the true (actual) values
- MAE is a measure of the **average magnitude of error** generated by the regression model
- The mean absolute error (MAE) is calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

- MAE is calculated by following these steps:
  1. Calculate the residual of every data point
  2. Calculate the absolute value (to get rid of the sign)
  3. Calculate the average of all residuals
- If MAE is zero, this indicates that the model predictions are perfect.

REGRESSION METRICS: MEAN SQUARE ERROR (MSE)

- Mean Square Error (MSE) is very similar to the Mean Absolute Error (MAE) but instead of using absolute values, squares of the difference between the model predictions and the training dataset (true values) is being calculated.
- MSE values are generally **larger** compared to the MAE since the **residuals are being squared**.
- In case of data outliers, MSE will become much larger compared to MAE
- In MSE, error increases in a **quadratic fashion** while the error increases in **proportional fashion in MAE**
- In MSE, since the error is being squared, any predicting error is being heavily penalized
- The MSE is calculated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- MSE is calculated by following these steps:
  1. Calculate the residual for every data point
  2. Calculate the squared value of the residuals
  3. Calculate the average of results from step #2

REGRESSION METRICS: ROOT MEAN SQUARE ERROR (RMSE)

- Root Mean Square Error (RMSE) represents the **standard deviation of the residuals** (i.e.: differences between the model predictions and the true values (training data)).
- RMSE can be **easily interpreted** compared to MSE because RMSE units match the units of the output.
- RMSE provides an estimate of how large the residuals are being dispersed.
- The RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$

- RMSE is calculated by following these steps:
  1. Calculate the residual for every data point
  2. Calculate the squared value of the residuals
  3. Calculate the average of the squared residuals
  4. Obtain the square root of the result

REGRESSION METRICS: MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

- MAE values can range from 0 to infinity which makes it difficult to interpret the result as compared to the training data.
- Mean Absolute Percentage Error (MAPE) is the equivalent to MAE but provides the error in a percentage form and therefore overcomes MAE limitations.
- MAPE might exhibit some limitations if the data point value is zero (since there is division operation involved)
- The MAPE is calculated as follows:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} |(y_i - \widehat{y}_i)/y_i|$$

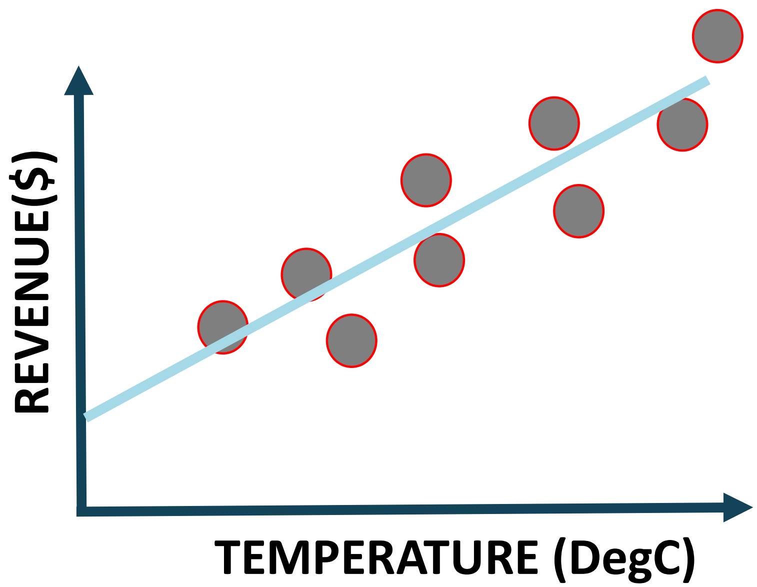
REGRESSION METRICS: MEAN PERCENTAGE ERROR (MPE)

- MPE is similar to MAPE but without the absolute operation
- MPE is useful to provide an insight of how many positive errors as compared to negative ones
- The MPE is calculated as follows:

$$MPE = \frac{100\%}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)/y_i$$

REGRESSION METRICS: R SQUARE (R^2)- COEFFICIENT OF DETERMINATION

- R-square or the coefficient of determination represents the proportion of variance (of y) that has been explained by the independent variables in the model.
- If $R^2 = 80$, this means that 80% of the increase in ice cream cart revenue is due to increase in temperature.
-



REGRESSION METRICS: R SQUARE (R^2)-COEFFICIENT OF DETERMINATION

- R-square or the coefficient of determination represents the proportion of variance (y) that has been explained by the independent variables (X) in the model.
- It provides an indication of goodness of fit and therefore a measure of how well unseen samples are likely to be predicted by the model, through the proportion of explained variance.
- Best possible score is 1.0
- A constant model that always predicts the expected value of y, disregarding the input features, would get a R^2 score of 0.0.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}$$

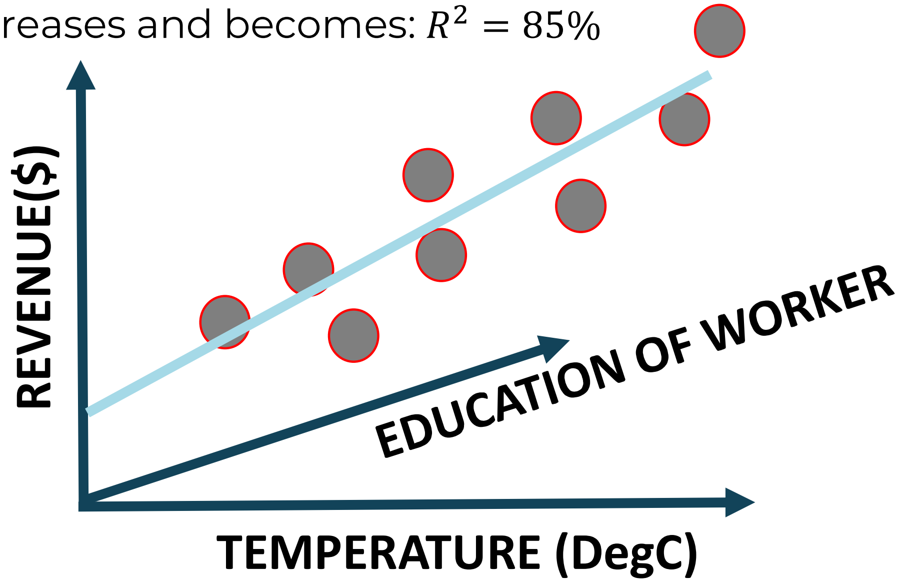
REGRESSION METRICS: R SQUARE (R^2)- COEFFICIENT OF DETERMINATION

- R-square represents the proportion of variance of the dependant variable (y) that has been explained by the independent variables.
- R-square provides an insight of goodness of fit.
- It gives a measure of how well unseen samples are likely to be predicted by the model, through the proportion of explained variance.
- Maximum R^2 value is 1
- A constant model that always predicts the expected value of y, disregarding the input features, will have an R^2 score of 0.0.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}$$

REGRESSION METRICS: ADJUSTED R SQUARE (R^2)-

- If $R^2 = 80$, this means that 80% of the increase in ice cream cart revenue is due to increase in temperature.
- Let's add another 'useless' independent variable, let's say level of education of worker to the Z-axis.
- Now R^2 increases and becomes: $R^2 = 85\%$



REGRESSION METRICS: ADJUSTED R SQUARE (R^2)-

- One limitation of R^2 is that it increases by adding independent variables to the model which is misleading since some added variables might be useless with minimal significance.
- Adjusted R^2 overcomes this issue by adding a penalty if we make an attempt to add independent variable that does not improve the model.
- Adjusted R^2 is a modified version of the R^2 and takes into account the number of predictors in the model.
- If useless predictors are added to the model, Adjusted R^2 will decrease
- If useful predictors are added to the model, Adjusted R^2 will increase
- K is the number of independent variables and n is the number of samples

$$R^2_{adjusted} = 1 - [\frac{(1 - R^2)(n - 1)}{n - k - 1}]$$