# Recap

- Classification
- Binary + Multi class classification
- Classification vs regression
- KNN
- Implementing KNN in scikit learn in python
- How to choose K (Grid search)
- Classification Metrics: Accuracy + misclassification rate
- Advantages and disadvantages of KNN

## Week 7: Data Science Part-Time Course

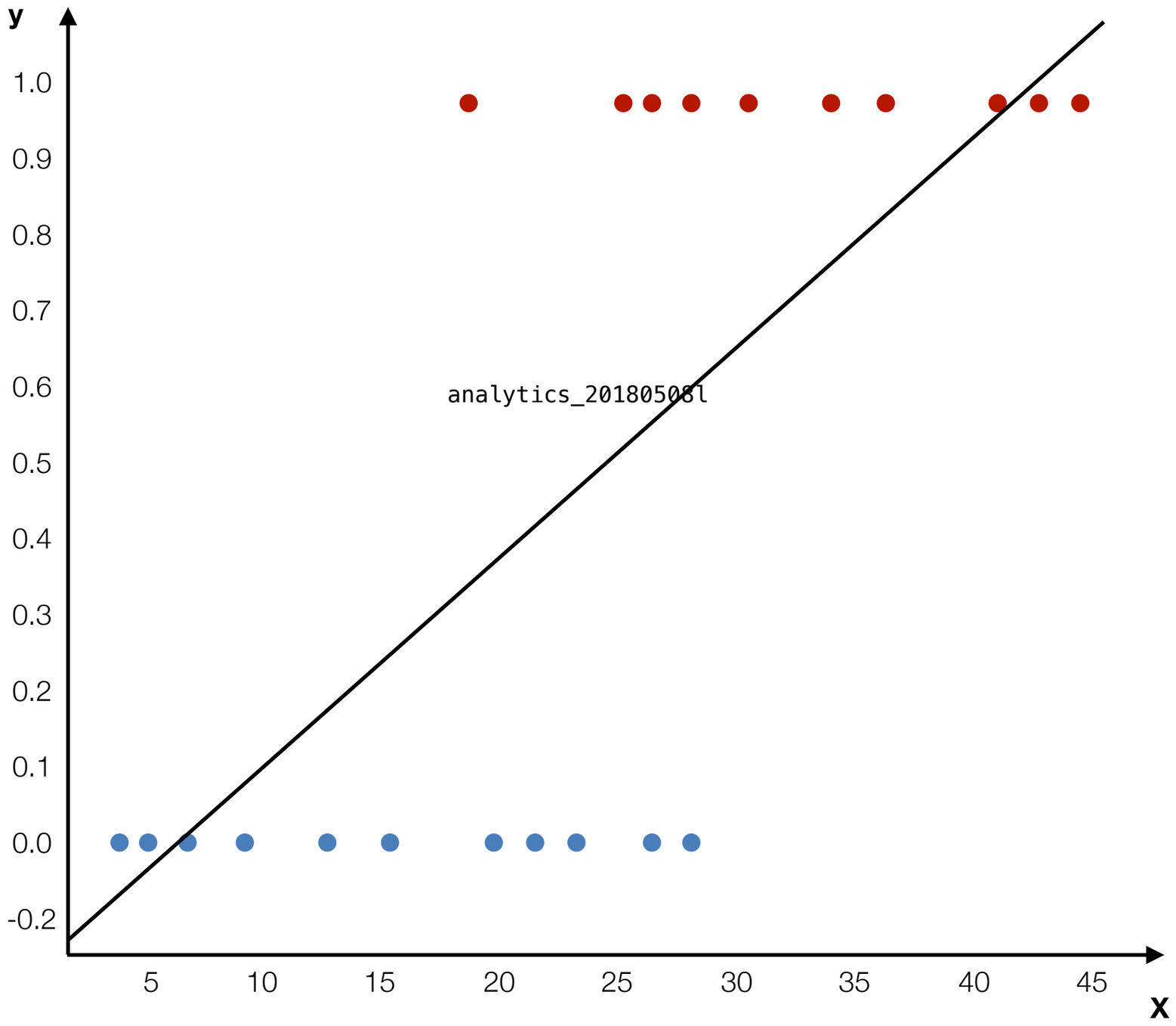
# Logistic Regression

Dami Lasisi

# Intro to Logistic Regression

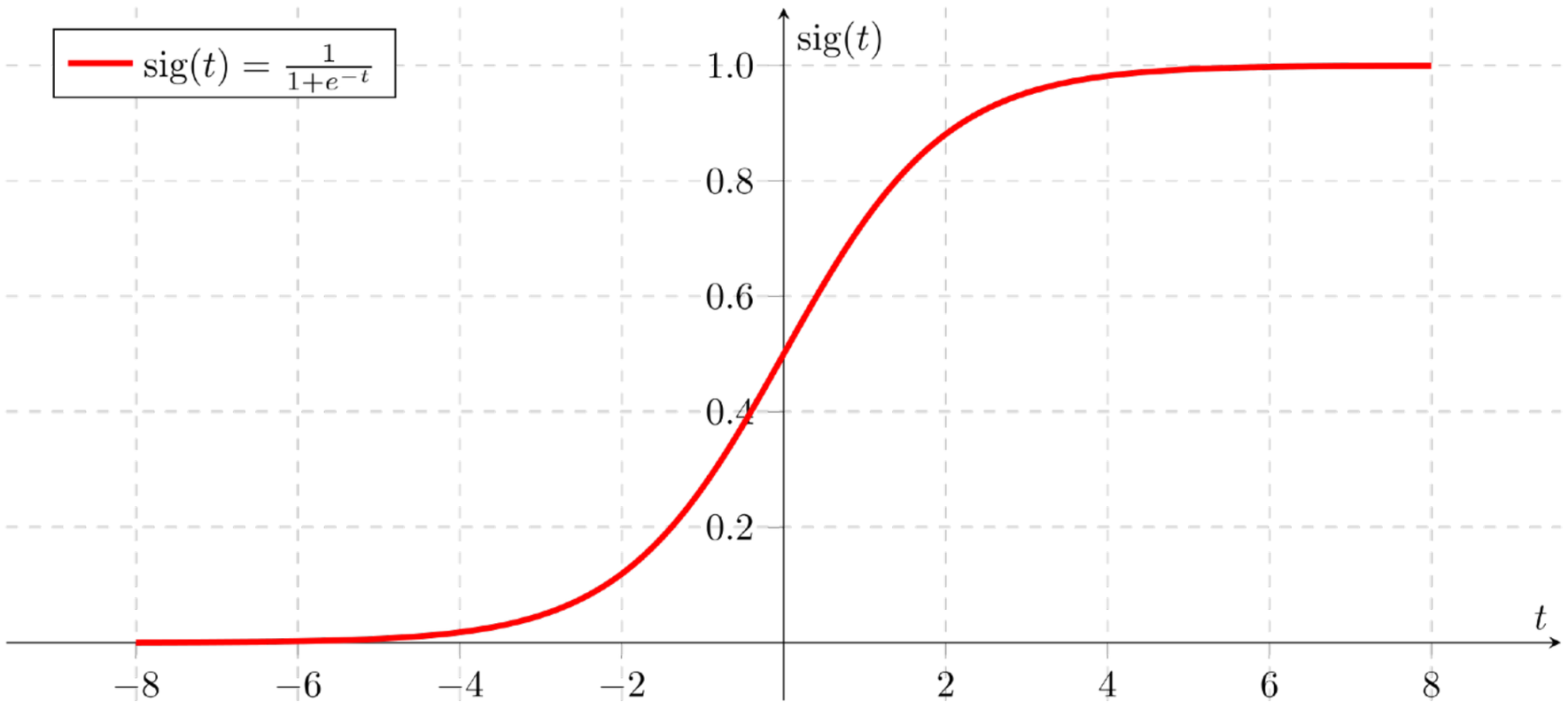
- **Finds the relationship between features and probability of a particular outcome**
- **Used when the dependent variable (target) is categorical**
- **Uses a linear approach to solve a classification problem while retaining the interpretability of a linear regression model**
- **Independent variables (predictors) can be categorical or continuous**

# Why not Linear Regression?



# The Sigma

$$y = b_0 + b_1 X_1 \longrightarrow p = P(y|X) = \frac{1}{1 + e^{-(b_0 + b_0 X)}}$$



# Log Odds Ratio

$$p = P(y|X) = \frac{1}{1 + e^{-(b_0 + b_1 X)}} \longrightarrow \log\left(\frac{p}{1 - p}\right) = b_0 + b_1 X_1$$

**Link function: inverse of the sigmoid (logistic) function**

**Example:**

$$\log\left(\frac{p_C}{1 - p_C}\right) = b_0 + 6.5\#cigarettesticks_1$$

# Maximum Likelihood Estimation (MLE)

- Used to obtain the model coefficients
- After the initial function has been estimated, the process is repeated until the log likelihood does not change significantly
- We find MLE using binomial distribution formula

# Classification Metrics

| Predicted Class \ True class | P | N |
|---|---|---|
| P | True Positive | False Positive |
| N | False Negative | True Negative |

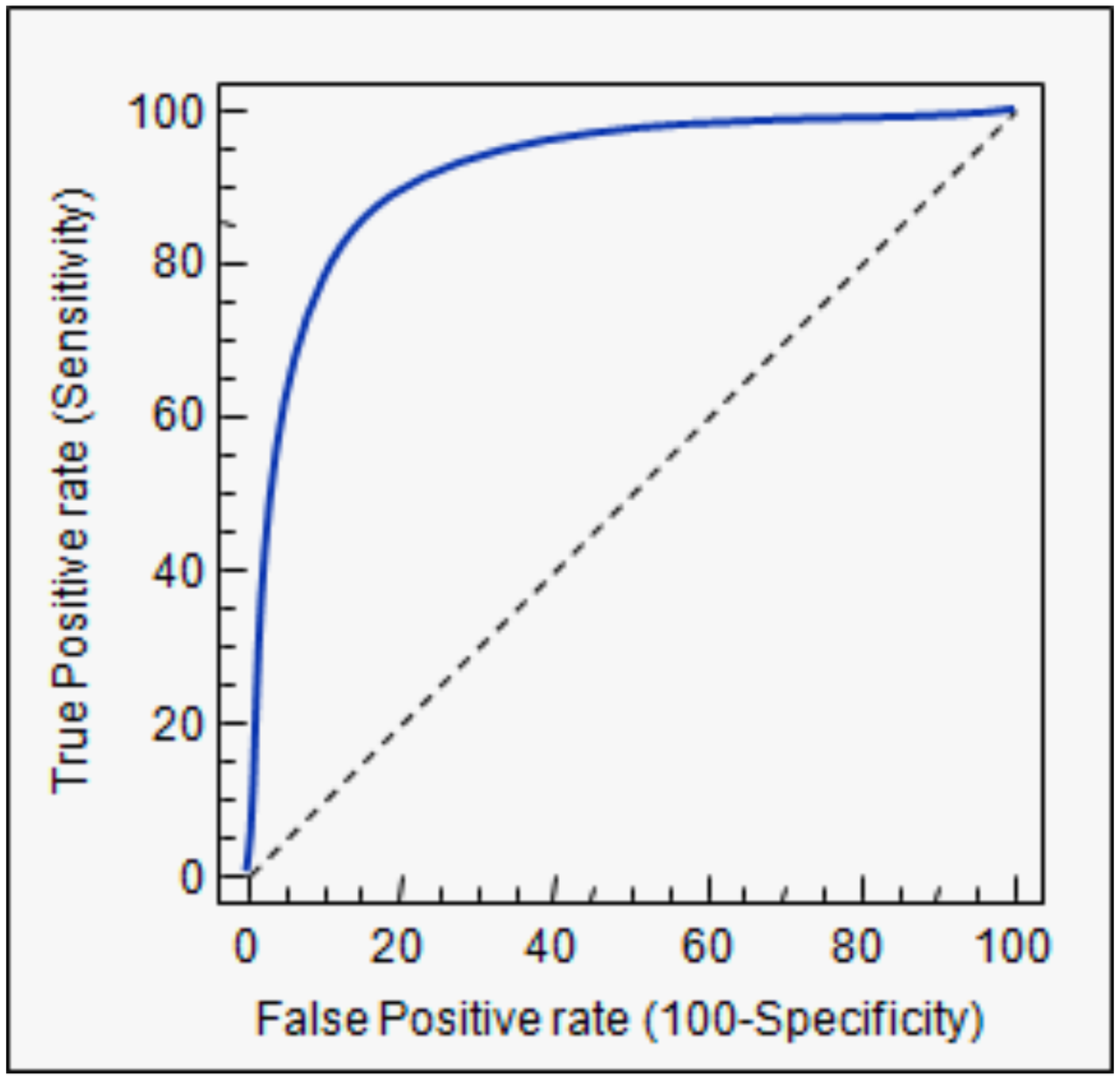
**TPR = TP/Actual positives**
**= TP/(TP+FN)**

**FPR = FP/Actual negative**
**= FP/(TN+FP)**
**= 1-TPR**

**TNR = TN/Actual negatives**
**= TN/(TN+FP)**

**FNR = FN/Actual positives**
**= FN/(TP+FN)**
**= 1-TNR**

# The Receiver Operating Curve (ROC)



# Advantage and Disadvantages of Logistic Regression

**Advantages:**

- Highly interpretable (if you remember how).
- Model training and prediction are fast.
- No tuning is required (excluding regularization).
- Features don't need scaling.
- Can perform well with a small number of observations.
- Outputs well-calibrated predicted probabilities.

**Disadvantages:**

- Presumes a linear relationship between the features and the log odds of the response.
- Performance is (generally) not competitive with the best supervised learning methods.
- Can't automatically learn feature interactions.