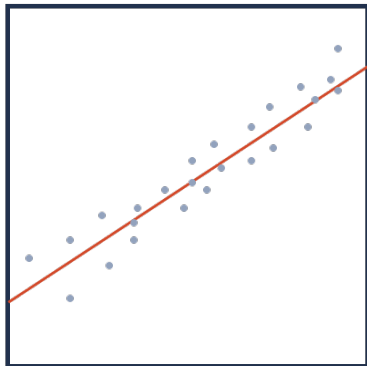
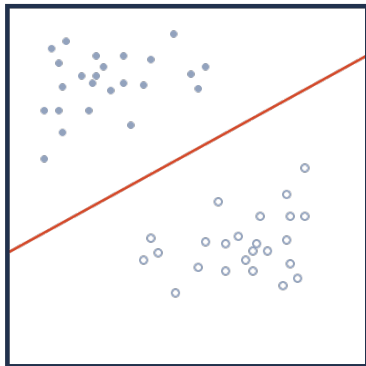


# Supervised Learning

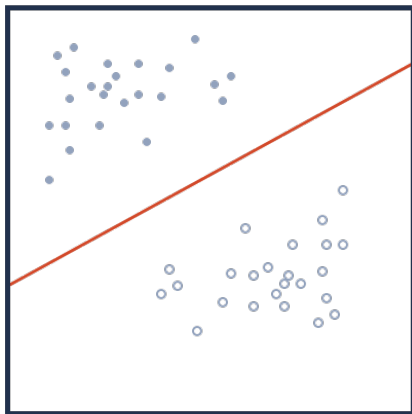
# Repeating the mistakes of the past



Regression



Classification



Classification

# Confusion Matrix

$c_{ij}$  is the count of data from class  $i$  that was labeled as class  $j$

$$\mathbf{C} = \begin{array}{c} \begin{array}{c} 1 \\ 2 \\ \vdots \\ n \end{array} \begin{array}{c} \text{Predicted} \\ \begin{array}{ccccc} 1 & 2 & 3 & \dots & n \end{array} \\ \left[ \begin{array}{ccccc} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ \vdots & \ddots & & & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \dots & c_{nn} \end{array} \right] \end{array} \begin{array}{c} \\ \\ \\ \text{Actual} \end{array} \end{array}$$

# Accuracy

$$\mathbf{C} = \begin{matrix} & \begin{matrix} \text{Predicted} \end{matrix} \\ \begin{matrix} \text{Actual} \end{matrix} & \begin{bmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ \vdots & \ddots & & & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \dots & c_{nn} \end{bmatrix} \end{matrix}$$

$$\text{Accuracy} = \frac{\sum_i c_{ii}}{\sum_i \sum_j c_{ij}}$$

# Accuracy

$$\mathbf{C} = \begin{matrix} & \begin{matrix} \text{Predicted} \end{matrix} \\ \begin{matrix} \text{Actual} \end{matrix} & \begin{bmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ \vdots & \ddots & & & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \dots & c_{nn} \end{bmatrix} \end{matrix}$$

$$\text{Accuracy} = \frac{\text{tr}(\mathbf{C})}{\sum_i \sum_j c_{ij}}$$

# True Positive, True Negative

$$\mathbf{C} = \begin{matrix} & \text{Predicted} \\ \begin{matrix} \text{Actual} \\ \left[ \begin{array}{ccccc} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ \vdots & \ddots & & & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \dots & c_{nn} \end{array} \right] \end{matrix} \end{matrix}$$

$$TP_i = c_{ii}$$

$$TN_i = \sum_{j \neq i, k \neq i} c_{jk}$$

# True Positive, True Negative

$$\mathbf{C} = \begin{matrix} & \text{Predicted} \\ \begin{matrix} \text{Actual} \\ \left[ \begin{array}{ccccc} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ \vdots & \ddots & & & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \dots & c_{nn} \end{array} \right] \end{matrix} \end{matrix}$$

$$TP_i = c_{ii}$$

$$TN_i = \sum_{j \neq i, k \neq i} c_{jk}$$



# False Positive, False Negative

$$\mathbf{C} = \begin{matrix} & \text{Predicted} \\ \begin{matrix} \text{Actual} \\ \left[ \begin{array}{ccccc} C_{11} & C_{12} & C_{13} & \dots & C_{1n} \\ C_{21} & C_{22} & C_{23} & \dots & C_{2n} \\ \vdots & \ddots & & & \vdots \\ C_{n1} & C_{n2} & C_{n3} & \dots & C_{nn} \end{array} \right] \end{matrix} \end{matrix}$$

$$FP_i = \sum_{j \neq i} C_{ji}$$

$$FN_i = \sum_{j \neq i} C_{ij}$$

# False Positive, False Negative

$$\mathbf{C} = \begin{matrix} & \text{Predicted} \\ \begin{matrix} \text{Actual} \\ \vdots \\ \end{matrix} & \begin{bmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ \vdots & \ddots & & & \vdots \\ c_{n1} & c_{n2} & c_{n3} & \dots & c_{nn} \end{bmatrix} \end{matrix}$$

$$FP_i = \sum_{j \neq i} c_{ji}$$

$$FN_i = \sum_{j \neq i} c_{ij}$$

# Summary

$$\mathbf{C} = \begin{matrix} & \begin{matrix} \text{Predicted} \\ C_{11} & C_{12} & C_{13} & \dots & C_{1n} \\ C_{21} & C_{22} & C_{23} & \dots & C_{2n} \\ \vdots & \ddots & & & \vdots \\ C_{n1} & C_{n2} & C_{n3} & \dots & C_{nn} \end{matrix} \\ \begin{matrix} \text{Actual} \\ \end{matrix} \end{matrix}$$

$$TP_i = c_{ii},$$

$$TN_i = \sum_{j \neq i, k \neq i} c_{jk},$$

$$FP_i = \sum_{j \neq i} c_{ji},$$

$$FN_i = \sum_{j \neq i} c_{ij}$$

$$TP_i = c_{ii},$$

$$FP_i = \sum_{j \neq i} c_{ji},$$

$$TN_i = \sum_{j \neq i, k \neq i} c_{jk},$$

$$FN_i = \sum_{j \neq i} c_{ij}$$

$$Pr_i = \frac{TP_i}{TP_i + FP_i}$$

$$TP_i = c_{ii},$$

$$FP_i = \sum_{j \neq i} c_{ji},$$

$$TN_i = \sum_{j \neq i, k \neq i} c_{jk},$$

$$FN_i = \sum_{j \neq i} c_{ij}$$

$$Re_i = \frac{TP_i}{TP_i + FN_i}$$

$$Pr_i = \frac{TP_i}{TP_i + FP_i}$$

$$Re_i = \frac{TP_i}{TP_i + FN_i}$$

$$F1_i = \frac{2 \times Pr_i \times Re_i}{Pr_i + Re_i}$$

# Class Metrics Summary

$$TP_i = c_{ii},$$

$$TN_i = \sum_{j \neq i, k \neq i} c_{jk},$$

$$Pr_i = \frac{TP_i}{TP_i + FP_i},$$

$$FP_i = \sum_{j \neq i} c_{ji},$$

$$FN_i = \sum_{j \neq i} c_{ij}$$

$$Re_i = \frac{TP_i}{TP_i + FN_i}$$

$$F1_i = \frac{2 \times Pr_i \times Re_i}{Pr_i + Re_i}$$

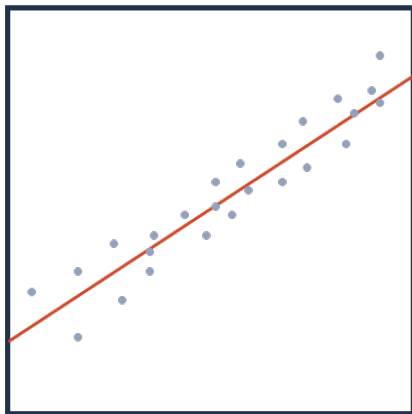
# Summary Metrics

$$Pr_{micro} = \frac{\sum_i TP_i}{\sum_i (TP_i + FP_i)}$$

$$Pr_{macro} = \frac{\sum_i Pr_i}{n}$$

$$Pr_{weighted} = \sum_i \left( \frac{c_i}{\sum_j c_j} Pr_i \right).$$





Regression

# Mean Squared Error

$$\text{MSE} = \frac{\sum_i (y_i - f_i)^2}{n}$$

$$\text{RMSE} = \sqrt{\frac{\sum_i (y_i - f_i)^2}{n}}$$

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \hat{y})^2},$$

# Methods

## Demo



**Chet Haase**

@chethaase



A Machine Learning algorithm walks into a bar.  
The bartender asks, "What'll you have?"  
The algorithm says, "What's everyone else having?"

7:24 AM - Nov 1, 2017

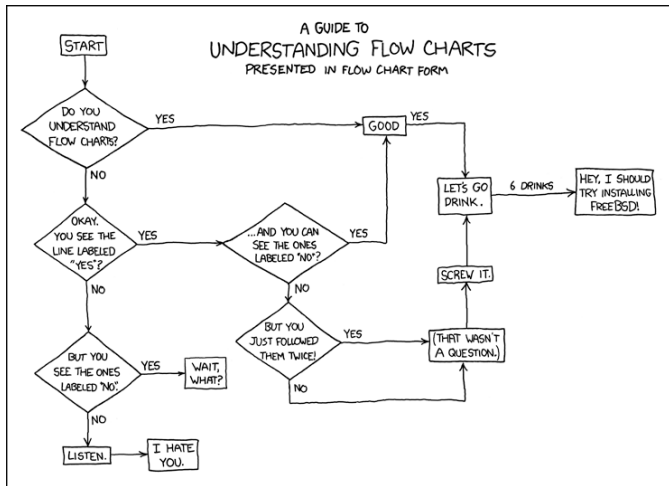
♡ 8,486 💬 4,342 people are talking about this



# $k$ Nearest Neighbors Considerations

- $k$ NN delays the computational effort until the inference/prediction stage
- $k$  is a hyperparameter that can affect the model's performance
- The way we calculate “nearest points” can significantly affect how our model performs. We can use an  $\ell_p$ -norm for our distance metric and  $\ell_1$ , and  $\ell_2$  are usually effective.

# Decision Trees



<https://xkcd.com/518/>



$$\text{Gini}_m = \sum_i p_{mi}(1 - p_{mi})$$

$$\text{Cross-entropy}_m = - \sum_i p_{mi} \log p_{mi}$$

# Decision Trees

- The splitting criterion or measure for impurity is a hyperparameter
- The stopping criterion for training is a critical hyperparameter to prevent overfitting of the data

# Random Forests

An ensemble of decision tree models.

# Random Forests Considerations

Same as decision trees plus,

- the number of decision trees used to create the random forest model is a hyperparameter to consider.

# Questions

These slides are designed for educational purposes, specifically the CSCI-470 Introduction to Machine Learning course at the Colorado School of Mines as part of the Department of Computer Science.

Some content in these slides are obtained from external sources and may be copyright sensitive. Copyright and all rights therein are retained by the respective authors or by other copyright holders. Distributing or reposting the whole or part of these slides not for academic use is HIGHLY prohibited, unless explicit permission from all copyright holders is granted.