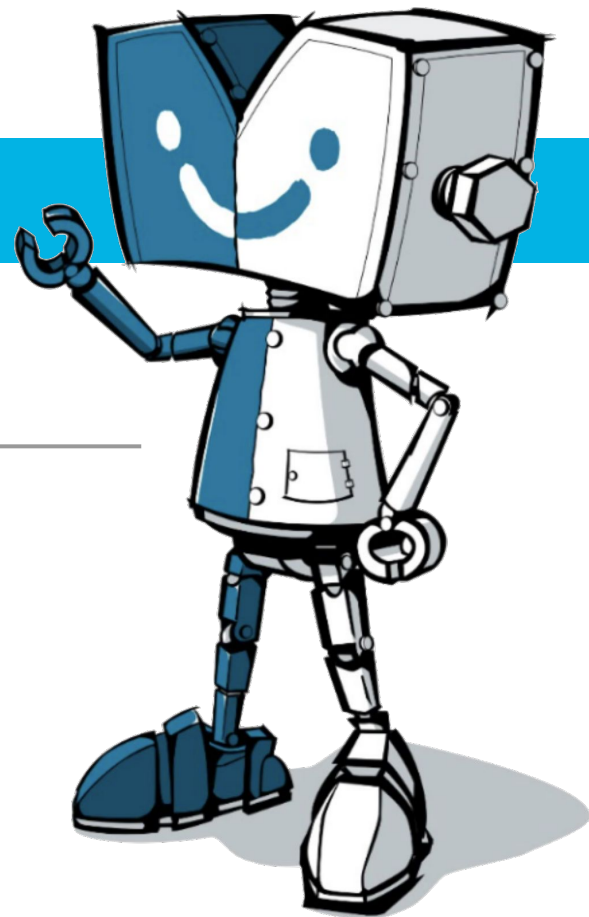
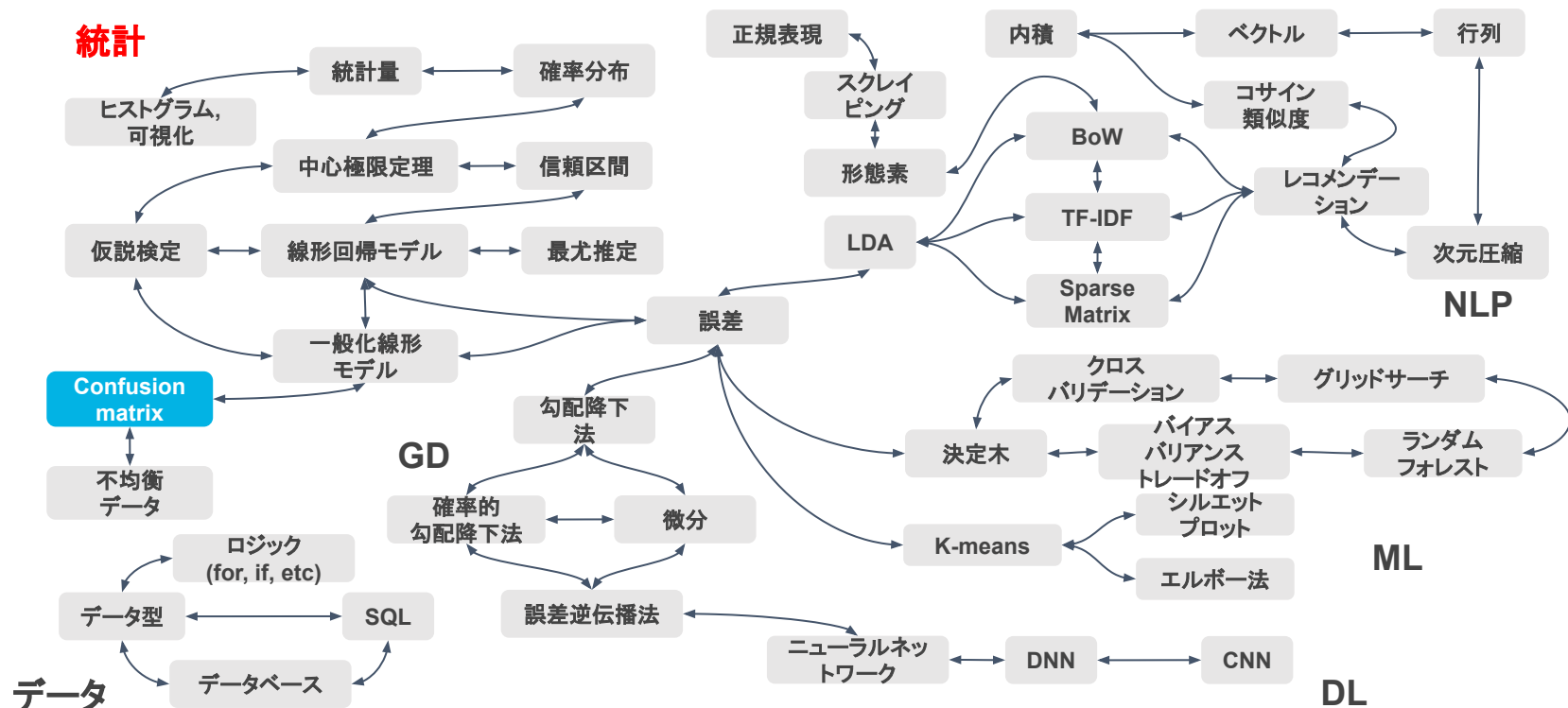




サンプルテキスト(混同行列)



Confusion matrix(混同行列)の位置づけ



小ケース | 緊急地震速報の例

- スマートフォンに内蔵されている、緊急地震速報の予測モデルの評価をすることとなりました。
- 予測と実際の結果は以下のような表であったとしましょう。
- あなたは、この結果をどのように評価しますか？

		予測されたクラス	
		地震だ！	地震ではない
実際のクラス	地震であった	1件	19件
	地震でない	0件	980件

混同行列 (Confusion matrix) とは？

混同行列とは実際のクラスと予測クラスを以下のように整理したもの

		予測されたクラス	
		Positive	Negative
実際のクラス	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

混同行列から計算できる4つの指標

- 正解率 (Accuracy)
- 再現率 (Recall)
- 適合率 (Precision)
- F値 (F1-measure)

正解率(Accuracy)

正解率とは、正しく分類できた割合

$$\text{正解率} = \frac{\text{True Positive} + \text{True Negative}}{\text{全データ数}}$$

		予測されたクラス	
		Positive	Negative
実際のクラス	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

適合率(Precision)

適合率とは、Positiveと予測したもののうち本当にPositiveだった割合

$$\text{適合率} = \frac{\text{True Positive}}{\text{予測クラスがPositive}}$$

		予測されたクラス	
		Positive	Negative
実際のクラス	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

再現率とは、実際のPositiveのうちPositiveだと予測できた割合

$$\text{再現率} = \frac{\text{True Positive}}{\text{実際のクラスがPositive}}$$

		予測されたクラス	
		Positive	Negative
実際のクラス	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

一般的に適合率(Precision)と再現率(Recall)はトレードオフになりやすいので、両方を気にしたい時に使うのが F-measure

$$F\text{値} = \frac{2 * \text{再現率} * \text{適合率}}{\text{再現率} + \text{適合率}}$$



データサイエンティストを目指すならデータミックス

Data has a better idea

説明会予約

データサイエンティスト
育成講座をみる