

Feature Selection

After this video you will be able to..

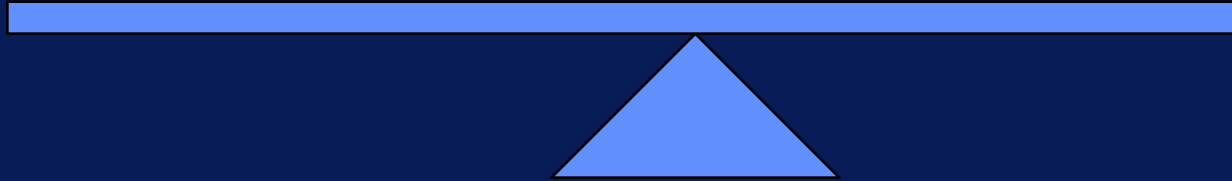
- Explain what feature selection involves
- Discuss the goal of feature selection
- List three approaches for selecting features

What is Feature Selection?

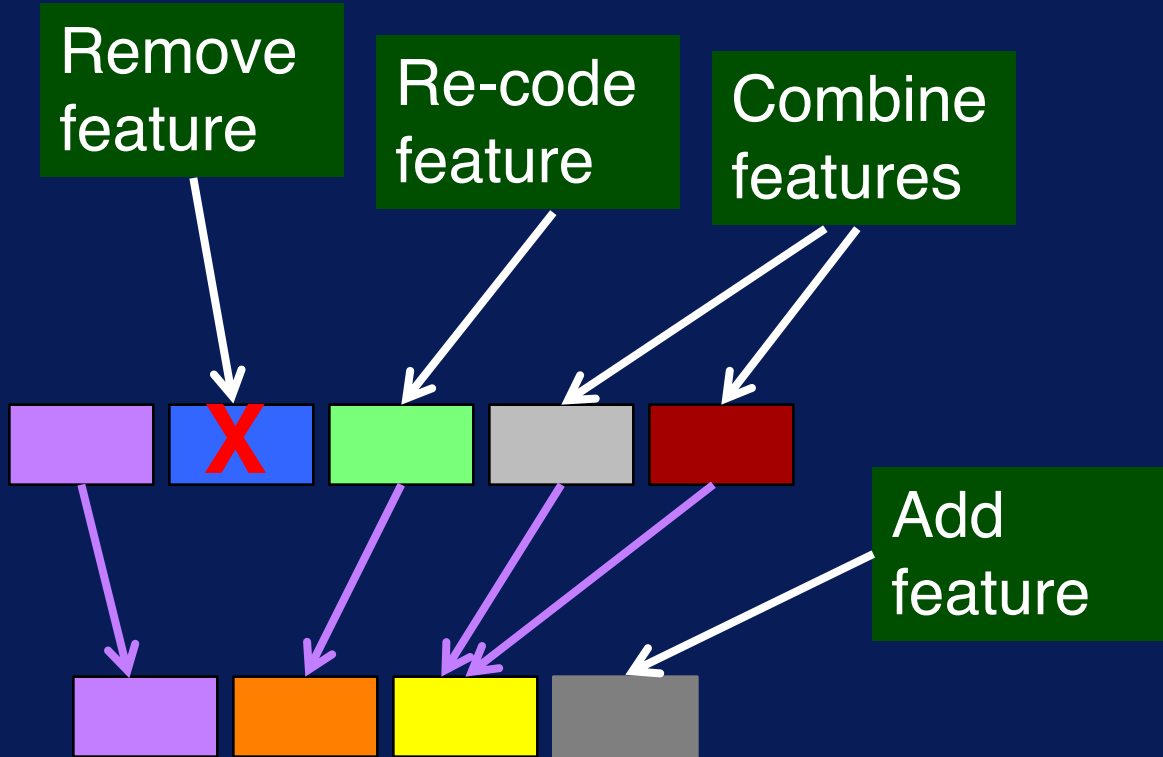
Characterize problem with
smallest set of features

Expressiveness

Size



Feature Selection Methods



Adding Features

New features derived from existing features

Name	State
Angela	AK
Sidney	CA
Ratan	WA
Kiril	OR
Zhou	CA



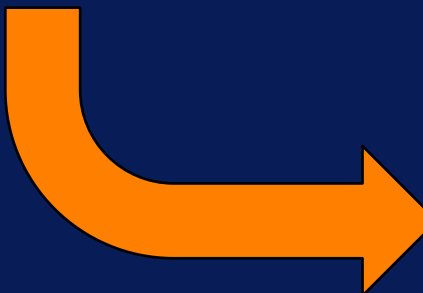
Name	State	<i>In-State</i>
Angela	AK	<i>F</i>
Sidney	CA	<i>T</i>
Ratan	WA	<i>F</i>
Kiril	OR	<i>F</i>
Zhou	CA	<i>T</i>

Removing Features

- Features that are very correlated
- Features with a lot of missing values
- Irrelevant features: ID, row number, etc.

Combining Features

Name	Height	Weight
Angela	1.8	68
Sidney	1.5	70
Ratan	2.0	84
Kiril	1.3	54
Zhou	2.0	61



Name	Height	Weight	<i>BMI</i>
Angela	180	68	21
Sidney	153	70	30
Ratan	204	84	20
Kiril	133	44	25
Zhou	208	81	19

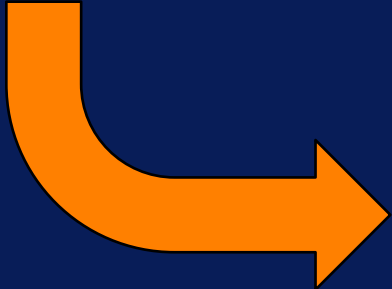
Recoding Features

- **Examples**
 - Discretization: re-format continuous feature as discrete
 - Customer's age => {teenager, young adult, adult, senior}

Breaking Up Features

Address
430 Park Drive, CA, 97283
7800 W. View Street, FL, 34642
1243 Mountain Ave., CO, 80334
1220 Mill Avenue, IL 54622
4345 Apple Lane, WA 98421

Address	State	Zip
430 Park Drive	CA	97283
7800 W. View Street	FL	34642
1243 Mountain Ave.	CO	80334
1220 Mill Avenue	IL	54622
4345 Apple Lane	WA	98421



Feature Selection Summary

- Goal: Select smallest set of features that best captures data for application.
- Domain knowledge is important
- aka 'feature engineering'