# Using Statistical Techniques for Feature Selection

**Janani Ravi**
CO-FOUNDER, LOONYCORN

www.loonycorn.com

# Overview

Methods for selecting and eliminating features

Variance thresholds

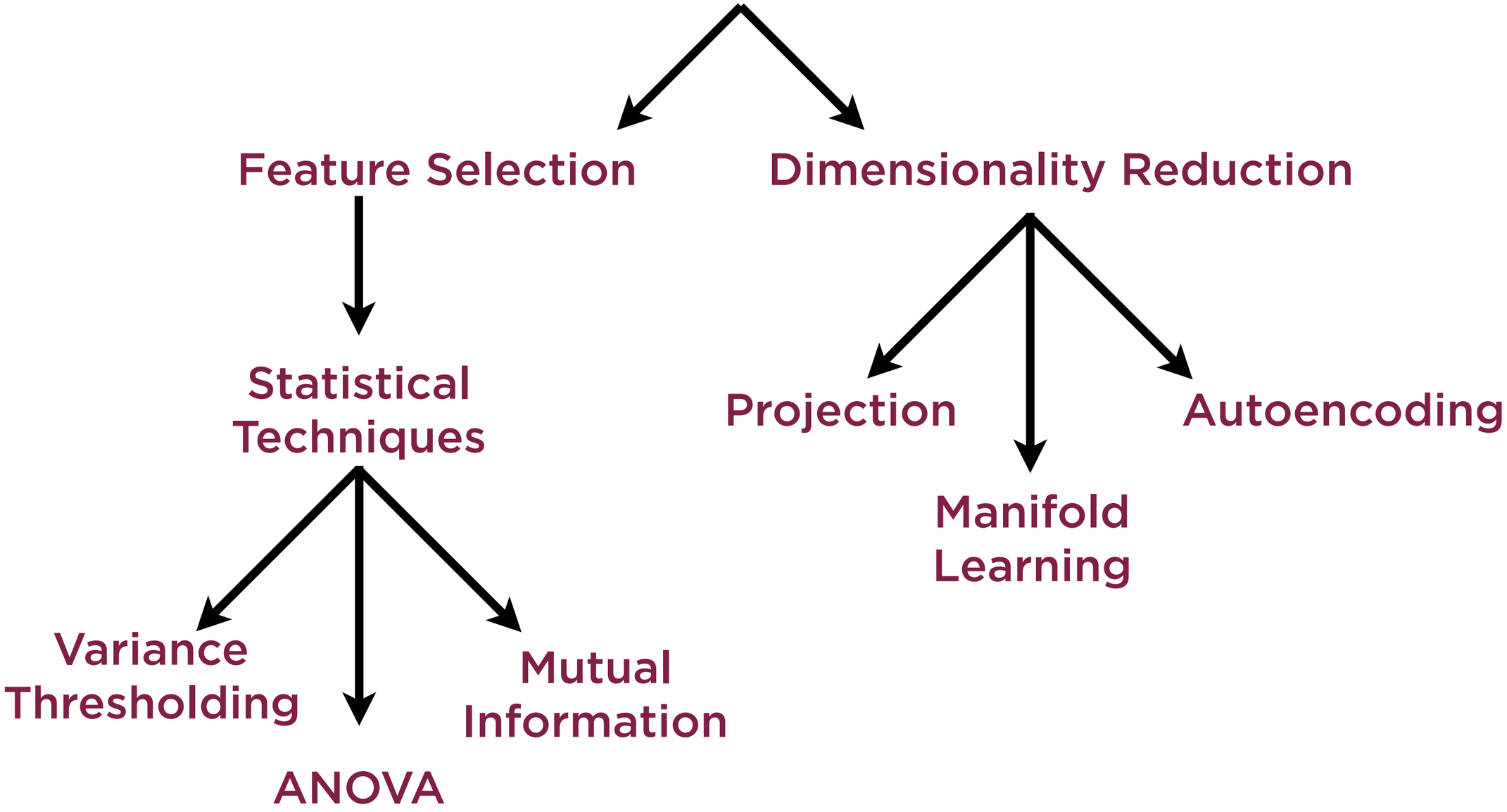
Univariate statistics

ANOVA, chi-square, mutual information

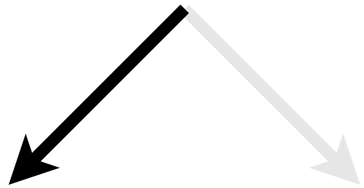
Dictionary learning and atom extraction

# Statistical Techniques for Feature Selection

# Reducing Complexity



# Reducing Complexity



**Feature Selection**

Choose a subset of original X variables

Dimensionality Reduction



Projection

Manifold Learning

Autoencoding

# Choosing Feature Selection

| Use Case | Possible Solution |
| --- | --- |
| Many X-variables | |
| Most of which contain little information | Feature selection |
| Some of which are very meaningful | |
| Meaningful variables are independent of each other | |

# Hypothesis

Proposed explanation for a phenomenon.

# Lady Tasting Tea

**Lady tasting tea: famous experiment**

**Was tea added before or after milk?**

**Muriel Bristol claimed she could tell**

# Lady Tasting Tea

**Null Hypothesis**

**($H_0$)**

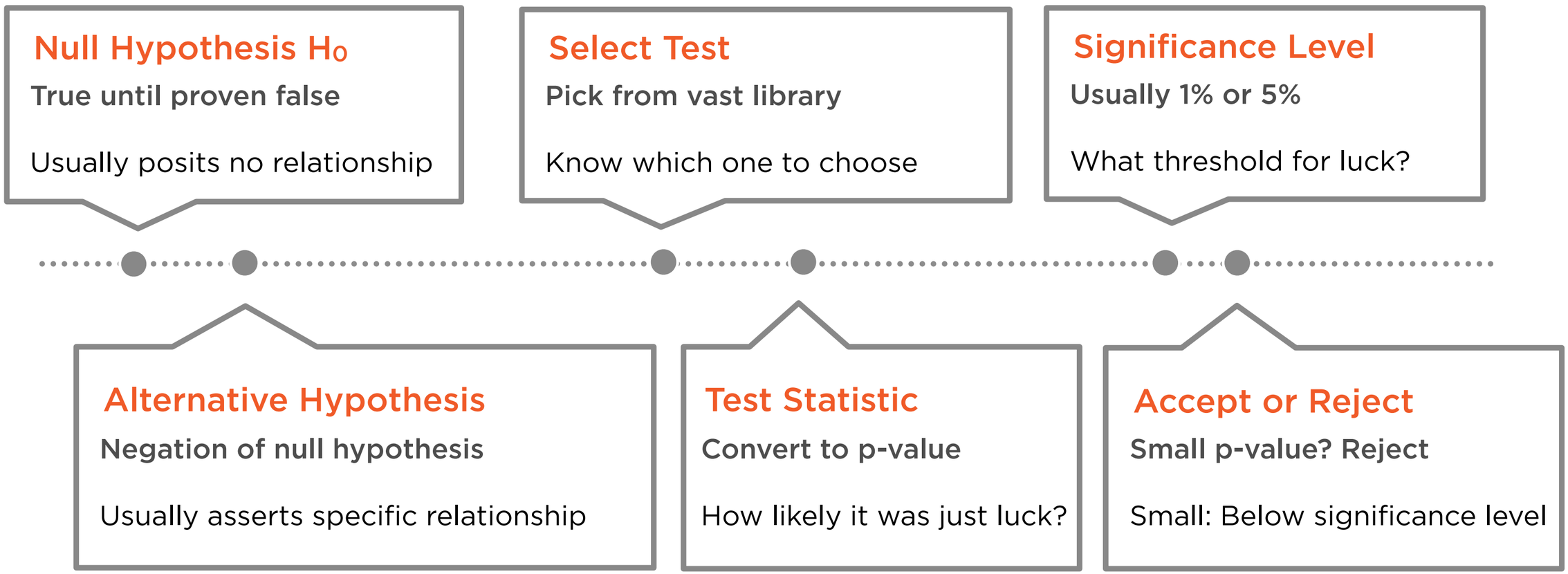
The lady cannot tell if milk was poured first

**Alternate Hypothesis**

**($H_1$)**

The lady can tell if milk was poured first

# Hypothesis Testing



# Statistical Techniques

**Variance Thresholding**

**Chi-square Test**

**ANOVA**

**Mutual Information**

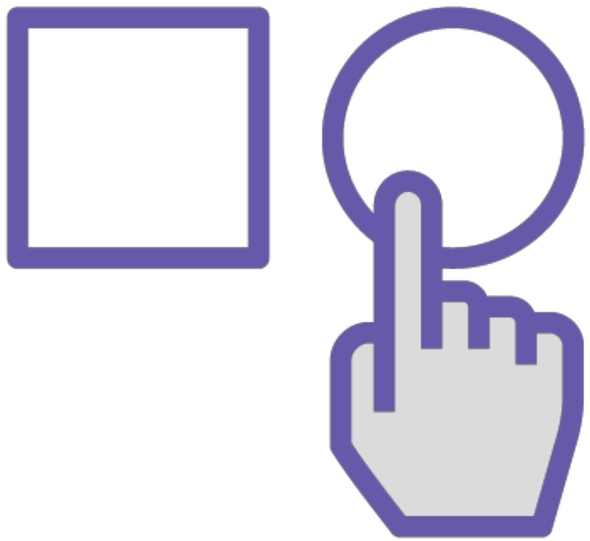
# Variance Thresholding

If all points have the same value for an X-variable, that variable adds no information. Extend this idea and drop columns with variance below a minimum threshold.

# Chi-square ($\chi^2$) Feature Selection

For each X-variable, use the Chi-square test to evaluate whether that variable and Y are independent. If yes, drop that feature. Used for categorical X and Y.

# Chi-square Feature Selection



**Does observed data deviate from those expected in a particular analysis?**

**Tests the effect of one variable on the outcome, univariate analysis**

**Sum of the squared difference between observed and expected data in all categories**

# ANOVA

***AN***alysis ***O***f ***VA***riance

# ANOVA

Looks across multiple groups of populations, compares their means to produce one score and one significance value

# ANOVA

Looks across multiple groups of populations, compares their means to produce one score and one significance value

# ANOVA Feature Selection

For each X-variable, use the ANOVA F-test to check whether mean of Y category varies for each distinct value of X. If not, drop that X-variable.

# Diabetes Risk

Underweight patients

Normal weight patients

Overweight patients

**Perform an ANOVA test to know whether the risk of diabetes is significantly different between these groups**

# ANOVA Hypotheses

**Null Hypothesis**

**($H_0$)**

$H_0$: All groups of patients are at an equal risk of diabetes

**Alternate Hypothesis**

**($H_1$)**

$H_0$: All groups of patients are NOT at an equal risk of diabetes

# F-statistic

$$F = \frac{\text{Variance between groups}}{\text{Variance within a group}}$$

# F-statistic

**If the groups are similar, F ~ 1**

**If the groups are different, F will be large**

# P-value

**Significance of the F-statistic**

**Smaller p-values indicate that the results are not due to chance**

**Large F-statistic and small p-value - means the null hypothesis can be rejected**

# Mutual Information

Measures the amount of information obtained on random variable by observing another.

# Mutual Information

Conceptually similar to using ANOVA F-test for feature selection; superior as it also captures non-linear dependencies (unlike ANOVA-based feature selection).

# Dictionary Learning

Representation learning method to find a sparse representation of input data.

# Demo

**Use variance threshold to select features for regression**

# Demo

**Use univariate statistics to select features for classification**

# Demo

**Using dictionary learning to find sparse representations of complex data**

# Summary

Methods for selecting and eliminating features

Variance thresholds

Univariate statistics

ANOVA, chi-square, mutual information

Dictionary learning and atom extraction