IS4242
INTELLIGENT SYSTEMS & TECHNIQUES

L2 – Pricing

Aditya Karanam

© Copyright National University of Singapore. All Rights Reserved.

Pricing

- What are the ways to improve profits of a company?
  - Profit = (Price - Cost)* Volume (or Sales Quantity)
- Assuming no decrease in price, improving unit volume by 1% yields a 3.3% increase in operating profit
- Assuming no loss of volume, a 1% improvement in price, increases operating profit by 11.1%
- The fastest and most effective for a company to realize its profits is by getting its pricing right
- What can be the right price for your product?
  - Simply, price above the cost of goods sold: cost-oriented pricing

Source: https://hbr.org/1992/09/managing-price-gaining-profit

Pricing Approaches: Cost-plus Pricing

- Cost-plus pricing:
  - Apply a pre-determined mark up to the cost in making or obtaining the product
    - E.g: Obtain total cost of production and add a mark up of 25% to obtain the price
  - Easy to estimate or measure
  - Easy to justify to various stakeholders
    - Costumers are generally willing to pay reasonable mark-up and investors have healthy profit margins
  - However, this strategy limits the orginization's ability to capture the customer's willingness to pay, which can be quite devastating!

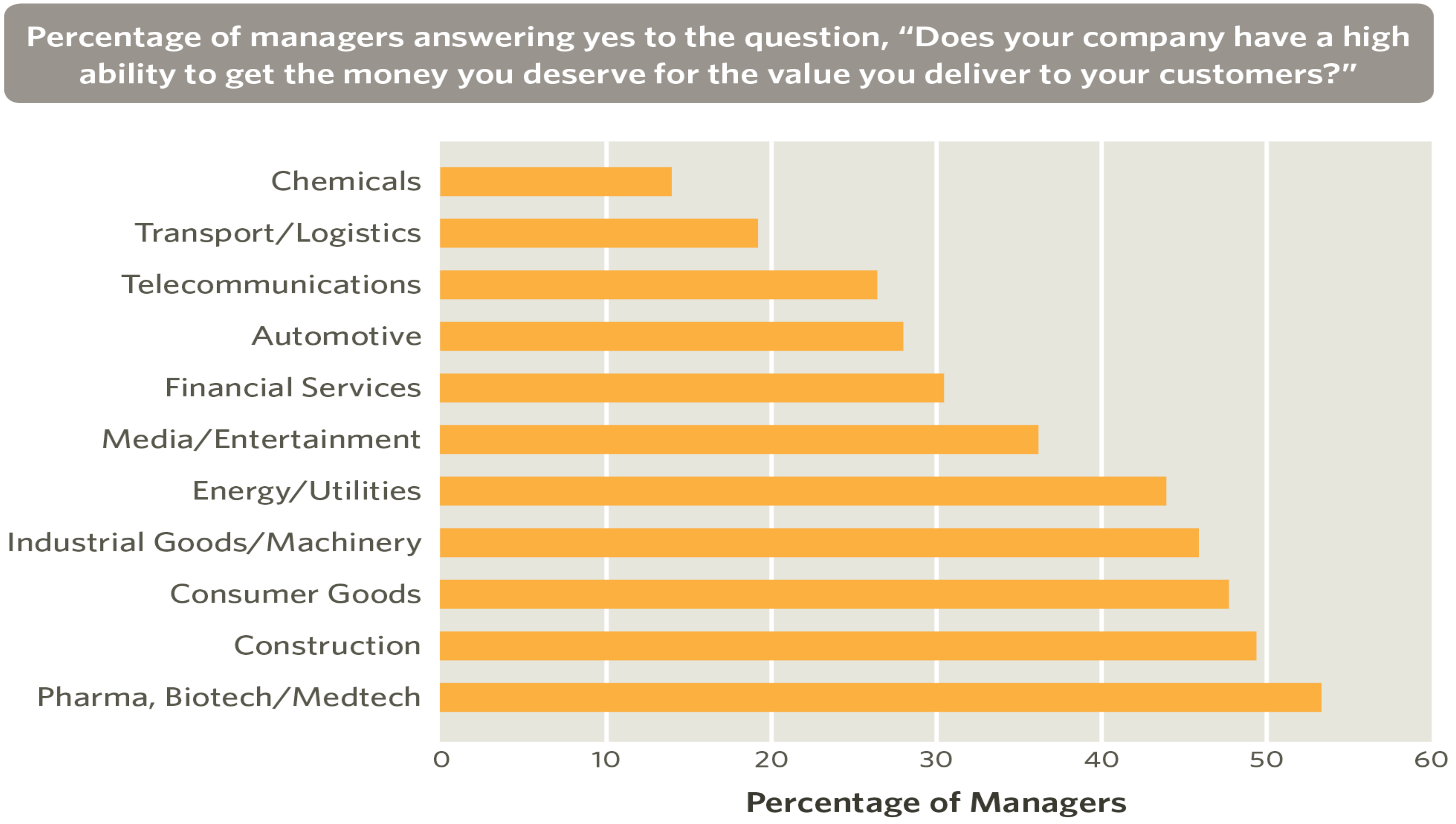
Pricing Approaches: Value Oriented Pricing

- Value Oriented Pricing:
  - Focus on the economic value of your product to the consumers (EVC)
  - Capture a portion of this value for the firm using the price

- Using value-oriented pricing firms earned 24% higher profits compared to industry peers
  - However, this strategy is challenging to implement

Value Oriented Pricing

- Survey on the potential to capture the value that the firm creates by industry:



Source: https://www.hbs.edu/faculty/Pages/item.aspx?num=45851

Value Oriented Pricing

- Economic Value to Consumer (EVC) argues that customer will buy the product only when its value outweighs the value of the next best alternative
  - Let *a* be the product you want to price and *b* be the next best alternative in the market
  - *Then:*

$$Value_a - Price_a \geq Value_b - Price_b$$

$$Price_a \leq Value_a - Value_b + Price_b$$

$$Price_a \leq Price_b + DifferentialValue_{ab}$$

- This price captures the maximum willingness to pay by the customer or their economic value
- To sell a product, a firm needs to price *at or below* its competitor's price plus the differential value its product provides to the consumer

https://ocw.mit.edu/courses/15-818-pricing-spring-2010/2318b493c551d7770340c79ff6de0c7a_MIT15_818S10_lec02.pdf

Practical Implementation of EVC

- Identify what benefit your product provides
  - *Benefit* not the feature of the product
    - Ex: Xerox charges per photocopy
- Identify the closest competitive offering and its price
  - Best alternative for Delta Airlines is American Airlines not trains or cars
- Identify potential sources of differentiation value
  - This follows from the benefit
- Measure how much value these create

Calculating Economic Value: Example

- Server Alternatives for Toy Company

	New Product	Next Best Alternative
Probability of System Crash	1% over one year	20% over one year
Cost of system crash	\$100,000	\$100,000
Hours of operation	2,500	2,500
Operating cost per hour	\$15	\$10
Price	To be determined	\$75,000

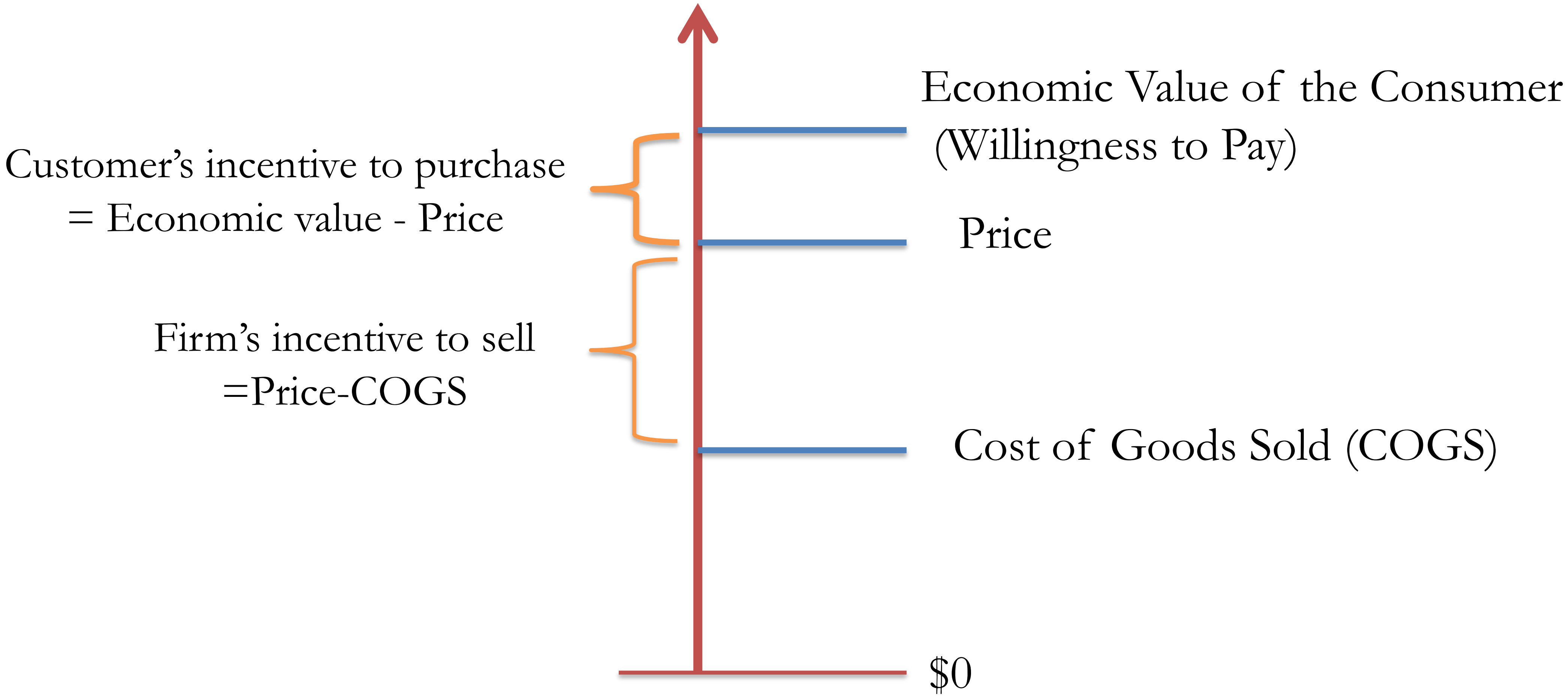
- EVC = Price of the alternative + value differential

  = \$75,000 + (20%*\$100000 – 1%*\$100000) -(\$15*2500 – \$10*2500)

  = \$75,000 + \$19,000 - \$12500 = \$81,500

Value Oriented Pricing Stick

Better off charging the price lower than the maximum price to have strong customer relations



https://sloanreview.mit.edu/article/why-the-highest-price-isnt-the-best-price/

https://online.hbs.edu/blog/post/value-based-strategy

Price Customization or Segmentation

- The economic value a product varies greatly across consumers based on several factors:
  - Tastes, Nature of use, Intensity of use, etc.
- Hence, the idea of uniform pricing is sub-optimal

- Price customization is achieved based on two major aspects:
  - Consumer Characteristics
  - Product Attributes

Consumer-Based Pricing: Implementation

- Use observable characteristics that correlate with the EVC
  - The characteristic used should clearly identify the group member
  - E.g.: demographic, gender, etc.

- Product must not be tradeable across group
  - Else, the products will be sold at a different price creating an alternate market

https://ocw.mit.edu/courses/15-818-pricing-spring-2010/24b2de274486a33ad373351ec961166f_MIT15_818S10_lec05.pdf

Consumer Based Pricing: Issues

- Price discrimination with anti-competitive intent
  - Lawful if the prices reflect different costs of dealing with different buyers or are the result of a seller's attempts to meet a competitors' offering

- Ethical issues with customer-based price discrimination
  - Must be culturally acceptable, else it may be perceived as illegal or unfair

- Irrespective of these cases customer-based pricing is important
  - In the early 2000s, 90% of those who suffer from AIDS could not afford the prices charged for AIDS drugs

Pricing Based on Product Attributes

- Design products such that customers signal their type (high or low value) through their choice of product
  - Ex: Incorporating a subscription price and authentication functionality in software products
- Requires you to find attributes that correlate with EVC
- Ensure that the integrity of the different products within the product line is maintained
  - Maintain fairness for low-segment users

How to obtain customized pricing?

- The main challenge in both cases is finding attributes that correlate with the EVC
  - Data: Surveys or historical data of consumers and product attributes
  - Regression techniques that help us to identify significant attributes
  - EVC could be difficult to obtain, we can use prices as proxy

Techniques For Pricing

© Copyright National University of Singapore. All Rights Reserved.

Application: Modeling House Prices

- A US-based company – Surprise Housing has ventured into Australian market
- Their strategy is to purchase houses at a price below their actual values and flip them on at a higher price
- For this purpose, the company has collected a data set from the sale of houses in Australia

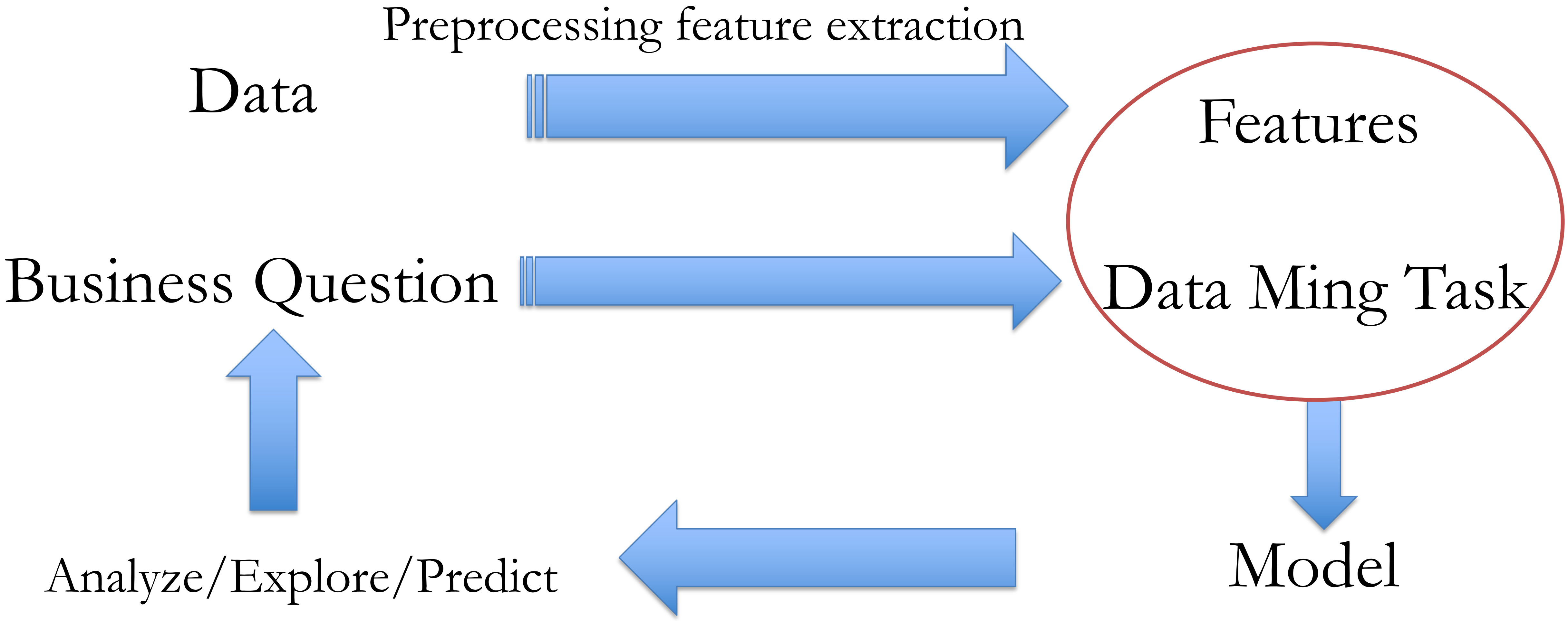
Data Description: Attributes

- Sale Price
- MSZoning: Identifies the general zoning classification of the sale
    - A:Agriculture, C:Commercial, …, RM:Residential Medium Density
- LotFrontage: Linear feet of street connected to property
- LotArea: Lot size in square feet
- Alley: Type of alley access to property: Gravl, Paved, etc.
- Utilities: Type of utilities available: AllPub, ELO: Electricity only
- OverallQual: Rates the overall material and finish of the house from 1-10
- ExterCond: Evaluates the present condition of the material on the exterior
- BsmtQual: Evaluates the height of the basement
- Other Attributes: Kitchen, fireplace, garage, etc.

Application: Modeling House Prices

- You are required to model the prices using available independent variables
- The management will use this model to understand how exactly the prices will vary with variables
- This model can be used to customize their prices accordingly to yield high return

Data Mining

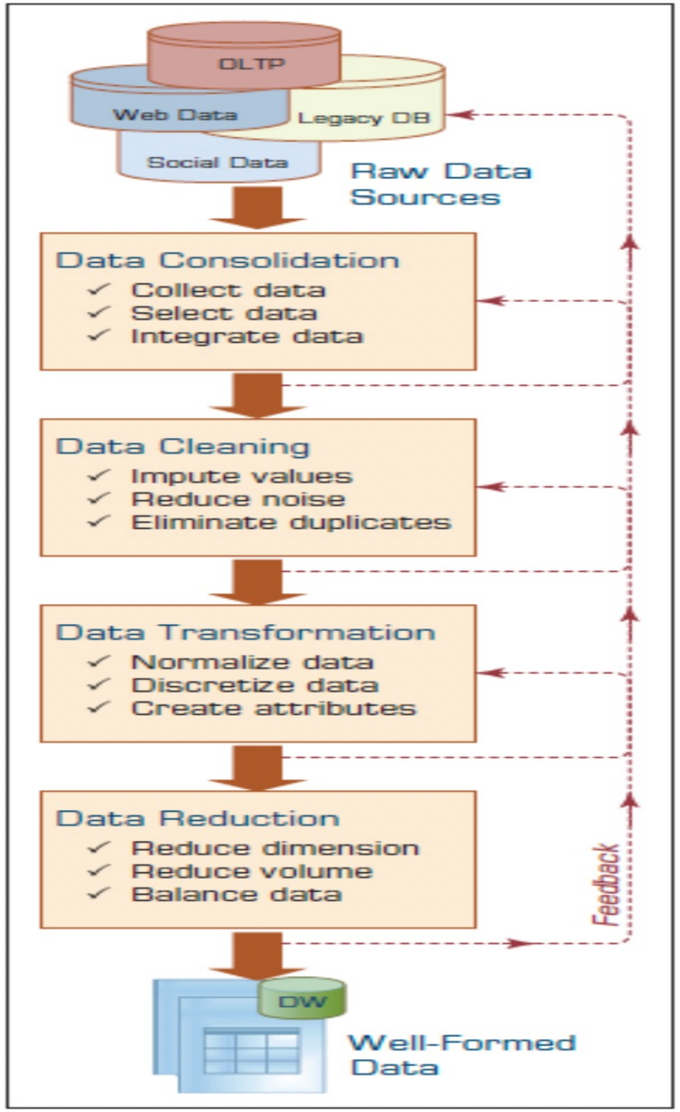

Preprocessing feature extraction
Data
Features
Business Question
Data Ming Task
Analyze/Explore/Predict
Model

Data Preprocessing

- Real world data is dirty, misaligned, overly complex and often inaccurate
  - Not ready for analytics

- Pre-processing:
  - Art - no clear methodology
  - Develops and improves with experience

Data Preprocessing



From Sharda et al., Business Intelligence, Analytics, and Data Science: A Managerial Perspective

Data Cleaning

- Missing Values
- Outliers
- Errors

Data Cleaning: Missing Values

- Missing Values
  - Do nothing
  - Remove columns or refrain from using them in the model
  - Remove Observations (if not many)
  - Impute with mean/median/mode/regression

	V1	V2	.	.	.	Vp
1						
2						
3						
.	?				?	
.						
.		?				
.						
.						
.		?				
.						
n						

- Outliers
- Errors

Data Cleaning: Missing Values

- Missing Values
- Outliers
  - Extreme or unrealistic values
  - Identify using data distribution
    - Ex: Standardization or Inter Quartile Range
  - Treat as a missing value or error
- Errors

	V1	V2	.	.	.	Vp
1						
2						
3						
.	?				?	
.						
.		?				
.						
.						
.		?				
.						
n						

Data Cleaning: Missing Values

- Missing Values
- Outliers
- Errors
  - Odd Values, inconsistent class labels, odd distributions
    - E.g.: Total Assets of a company is negative
  - Use domain expertise to correct or remove

	V1	V2	.	.	.	Vp
1						
2						
3						
.	?				?	
.						
.		?				
.						
.						
.		?				
.						
n						

Data Transformation

- Scaling
- Aggregation/Discretization/Binning
- Construct new variables

Data Transformation

- Scaling
  - Brining variables to same scale or range
    - In range [0, 1]: $\frac{V - V_{min}}{V_{max} - V_{min}}$
  - Normalization or standardization
    - $\frac{V - \mu_v}{\sigma_v}$ μ_v: Variable mean, σ_v: Standard deviation of the variable

- Aggregation/Discretization/Binning
- Construct new variables

	V1	V2	.	.	.	Vp
1	100	0.01				
2	150	0.03				
3	175	0.05				
.	180	0.01			?	
.						
.		?				
.						
.						
.		?				
n						

Data Transformation

- Scaling
- Aggregation/Discretization/Binning
  - Convert Numerical to categorical
  - Reduce categories using hierarchy/intervals

- Construct new variables

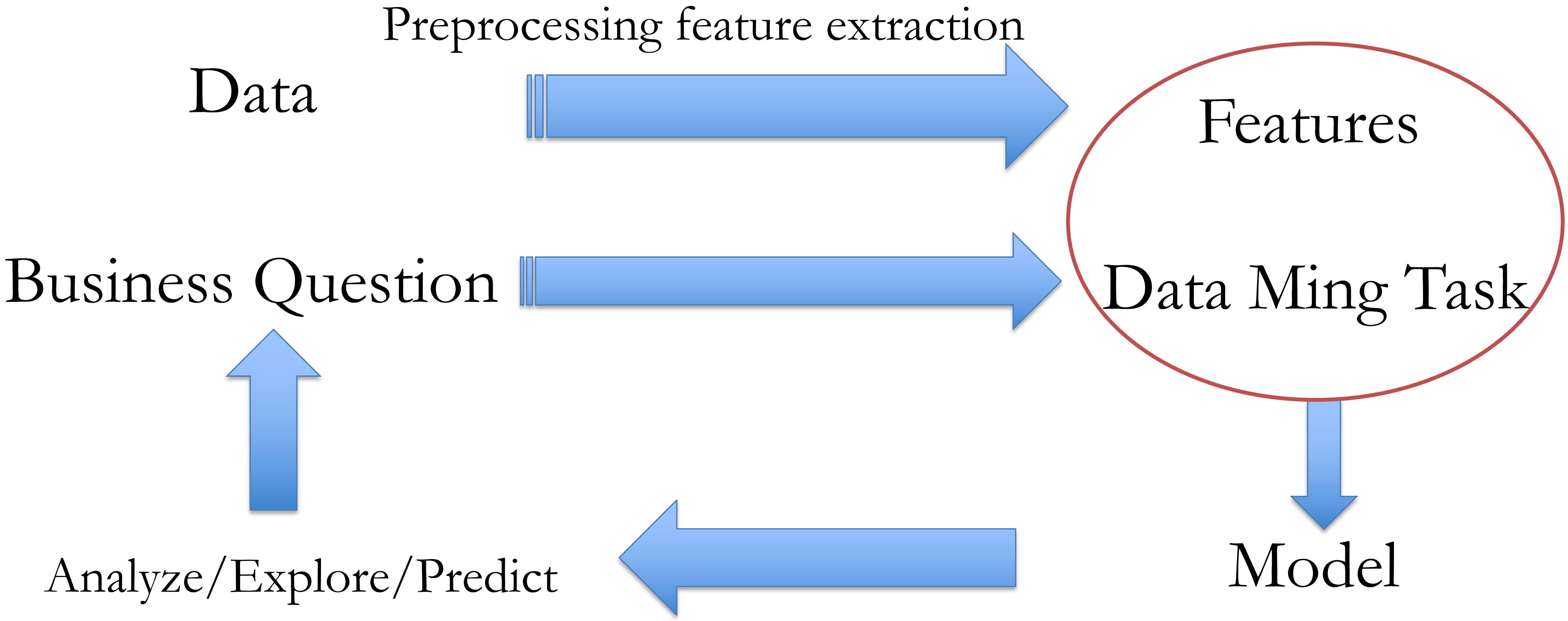
	V1	V2	.	.	.	Vp
1	100	0.01	A			
2	150	0.03	B			
3	175	0.05	C			
.	180	0.01	D		?	
.						
.						
.						
.						
.						
n						

Data Transformation

- Scaling
- Aggregation/Discretization/Binning
- Construct new variables
  - Addition, multiplication, log-transformation
  - One-hot encoding or creating dummy variable

	V1	V2	.	Va	Vb	Vp
1	100	0.01	A	1	0	
2	150	0.03	B	0	1	
3	175	0.05	C	0	0	
.	180	0.01	D	0	0	
.						
.						
.						
.						
.						
n						

Data Mining


Preprocessing feature extraction
Data
Features
Business Question
Data Ming Task
Model
Analyze/Explore/Predict

Statistical Modeling

- Task: Understand the relation between house attributes and price
- We want to compare the relation between two variables (house attributes and price)
  - Simply check the covariance or correlation between each variable and the price
  - Correlation only signals the direction of relationship, but does not help us in prediction
- Better way: Model it as a linear regression problem

Multiple (or Multivariate) Linear Regression

- Quantitative Response: Y
- Predictor variables: $X_1, X_2, \ldots, X_p$: $\mathbf{X}$
- Regression Y on $\mathbf{X}$ (predictor variables):
  - $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ \ldots + \beta_p X_p$
  - $Y = \boldsymbol{\beta X}$

Multiple Linear Regression

- Model Parameters: $\beta_0, \beta_1, .. \beta_p$
- Estimated from the data: $\widehat{\beta_0}, \widehat{\beta_1}, \dots, \widehat{\beta_p}$
- Prediction: $\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1} X_1 + \widehat{\beta_2} X_2 + \ \dots + \widehat{\beta_p} X_p$
- How to obtain coefficient estimates?

Finding the Estimates

- Start with some initial guess and move forward by minimizing error
  - $\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X_1 + \widehat{\beta_2}X_2 + \ \dots + \widehat{\beta_p}X_p$
- For *i*-th observation, residual or error: $e_i = y_i - \widehat{y_i}$
- Obtain residual sum of squares (RSS):

$$RSS = e_1{}^2 + e_2{}^2 + \ldots + e_n{}^2$$

- Least Squares: Find $\widehat{\beta_0}, \widehat{\beta_1}, \dots, \widehat{\beta_p}$ that minimizes RSS

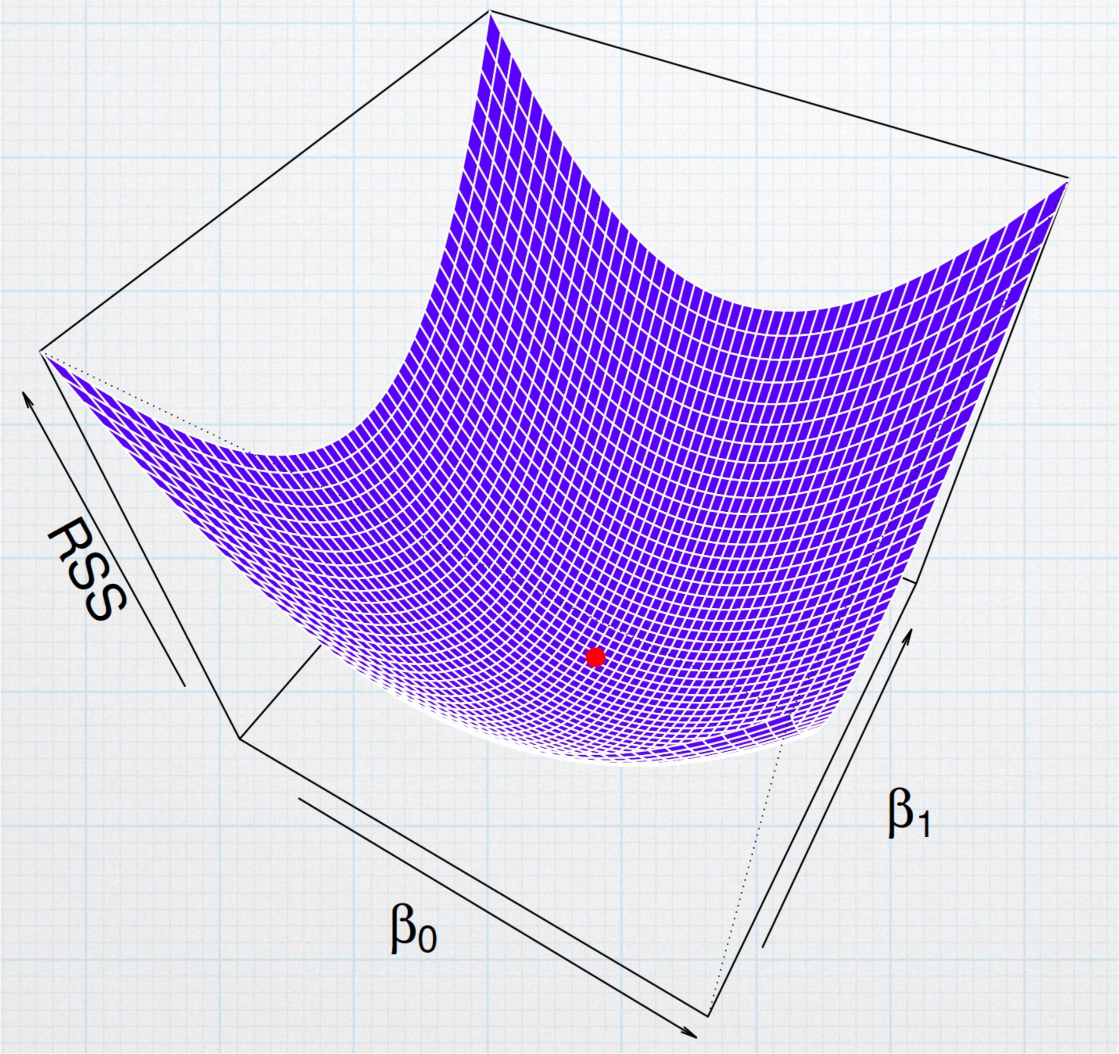
Finding the Estimates: Decision Surface

- Example: With one variable
  - $\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1} X_1$
  - $RSS = \sum_{i=1}^{n} {e_i}^2$
    - $e_i = y_i - \widehat{y_i}$

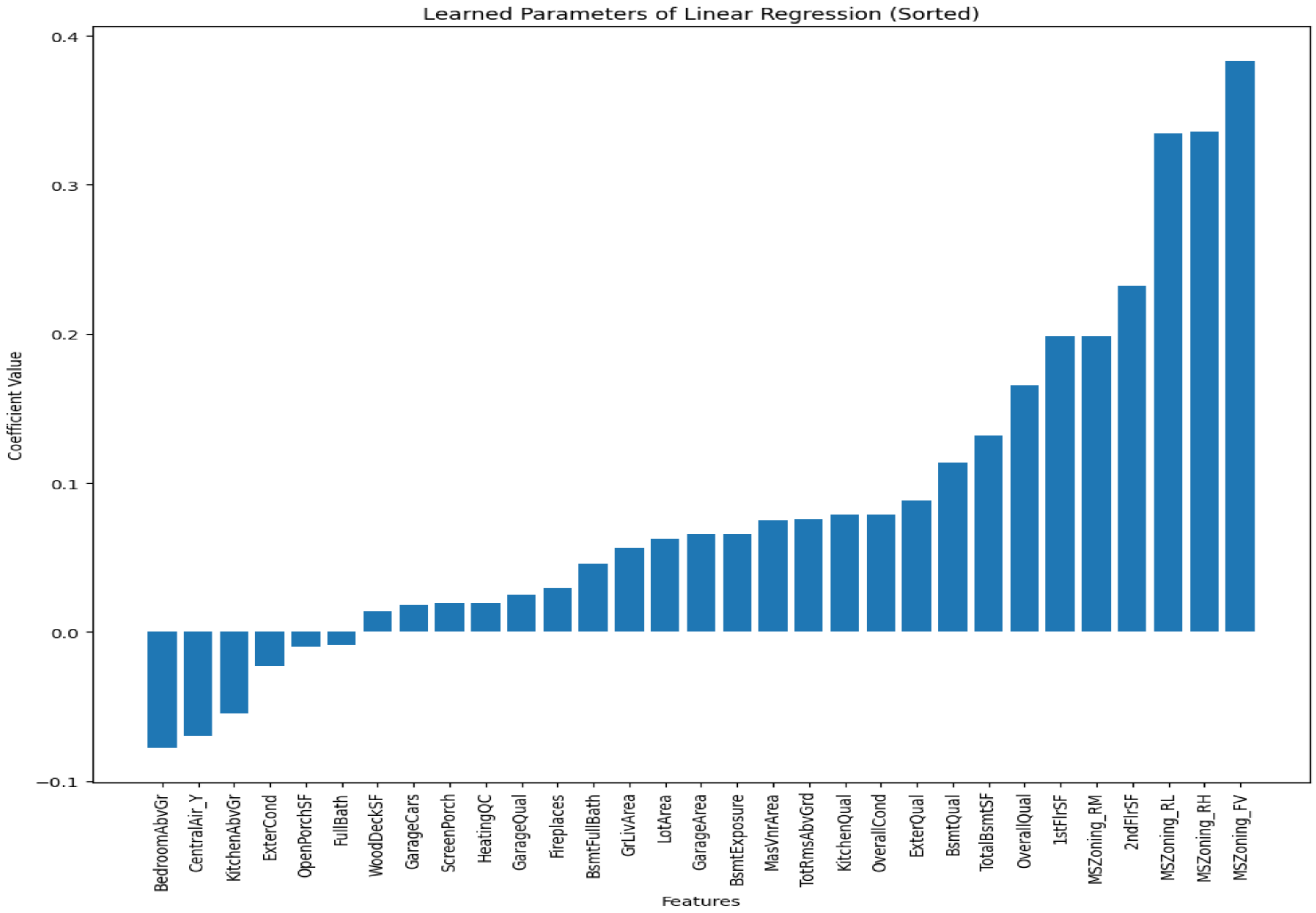
```python
import statsmodels.api as sm

# Added constant to X for the intercept
X = sm.add_constant(X)

# Fit the linear regression model
model = sm.OLS(y, X).fit()
model.summary()
```



Coefficient Estimates



Model Fit

- R-squared: proportion of variability in Y that can be explained using X
  - $R - Squared = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$
    - TSS: total variance in Y (before regression)
    - RSS: variance left unexplained (after regression)
  - ~1: Good fit, large proportion explained
  - ~0: Regression did not explain much variability
    - Linear model wrong and/or inherent error high
- R2-Squared in our example: 0.857
  - Which variables have significant impact? Hypothesis Testing

Hypotheses Testing in Linear Regression

- Null Hypothesis:
  - There is no relationship between X_i and Y
  - $H_0: \beta_i = 0$
- Alternate Hypothesis:
  - There is a non-zero relationship between X_i and Y
    - $H_1: \beta_i \neq 0$
- Hypothesis test: using *p*-values of the *t*-statistic

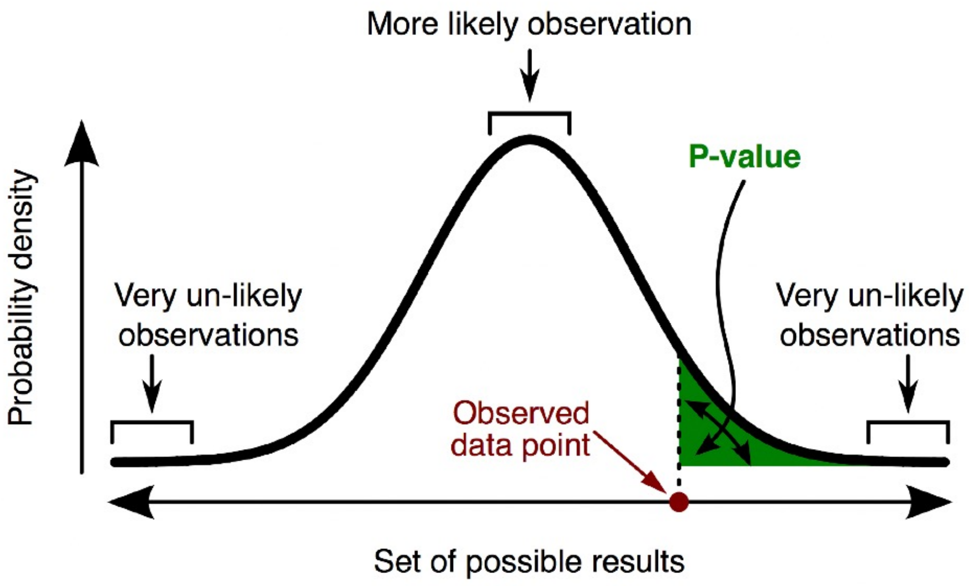
Hypothesis Test

- T-statistic of coefficient: $\frac{(Sample\ Coefficient\ - Hypothesised\ Coefficient)}{SE(Coefficient)}$
- For simple linear regression: $\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X_1$
  - $t = \frac{\widehat{\beta_1} - 0}{SE(\widehat{\beta_1})} = \frac{\widehat{\beta_1}}{SE(\widehat{\beta_1})}$
    - Variance of $\widehat{\beta_1} = \frac{RSS}{n-2}\left[\frac{1}{\sum_{k=1}^{n}(x_i - \bar{x})^2}\right]$

- Is our estimate($\widehat{\beta_1}$) sufficiently far from zero to confidently reject the null hypothesis?
  - If SE is low, then smaller non-zero estimates may be sufficient
  - If SE is high, then larger estimates required

P-value

- Probability, *under the null hypothesis,* of observing a value of the test statistic that is same as or more extreme than what was actually observed
  - A measure of evidence against the null hypothesis
  - Smaller the p-value, stronger is the evidence.
    - $P < 0.1$ or 0.05 is used



P-Values of the coefficients

```
print("P-values for coefficients:")
print(model.pvalues)
```

[46] ✓ 0.0s

```
P-values for coefficients:
const            5.111732e-02
LotArea          3.147398e-07
OverallQual      7.169965e-17
OverallCond      6.865579e-11
MasVnrArea       1.162264e-10
ExterQual        5.525251e-07
ExterCond        3.844129e-02
BsmtQual         3.249227e-12
BsmtExposure     1.869802e-08
TotalBsmtSF      1.253316e-11
HeatingQC        1.258589e-01
1stFlrSF         1.722300e-02
2ndFlrSF         1.465454e-02
GrLivArea        6.112155e-01
BsmtFullBath     5.757695e-05
FullBath         5.916593e-01
BedroomAbvGr     7.251462e-07
KitchenAbvGr     5.863971e-06
KitchenQual      8.496024e-07
TotRmsAbvGrd     6.077111e-04
Fireplaces       1.795845e-02
GarageCars       4.478288e-01
GarageArea       4.979587e-03
GarageQual       1.713422e-02
...
MSZoning_RL      8.621630e-03
MSZoning_RM      1.206461e-01
CentralAir_Y     1.385401e-01
dtype: float64
```

- Almost all of the coefficients are significant
- Model is complex
- Not a good model in identifying the variables that make an impact on the sales

Feature Selection in Linear Regression

- If there are '*p*' features, then possible subsets of predictors: 2^p
  - Identify best subset of features based on R-square
- The complexity of feature selection increases exponentially with number of variables
- Alternate: Regression that penalize these coefficient values or regularizes the model complexity
  - Least Absolute Shrinkage and Selection Operator(LASSO) Regression

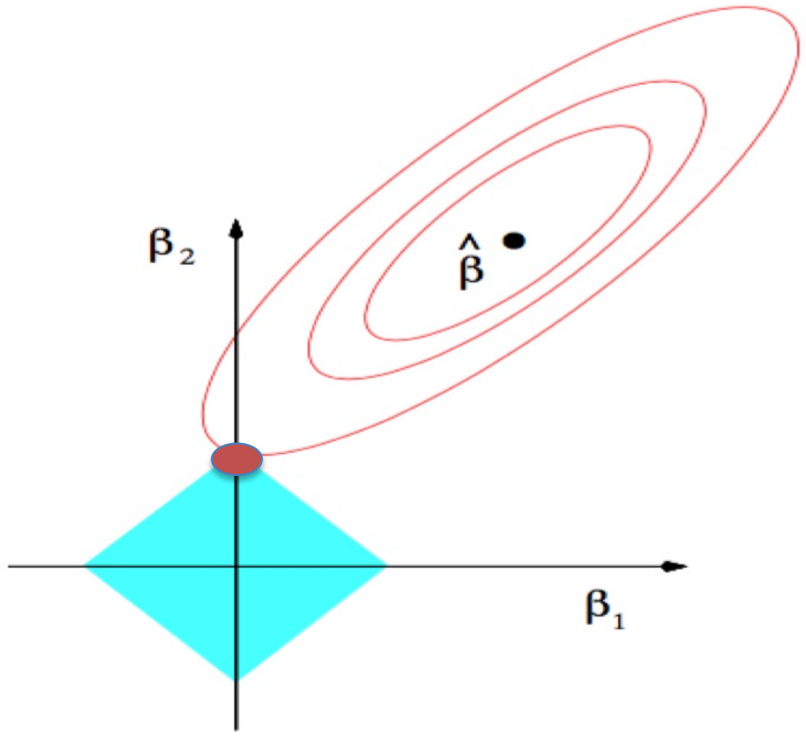
Least Absolute Shrinkage and Selection Operator (LASSO)

- Least Square Model with an additional penalty term based on the absolute values of the coefficients.
  - This term helps in shrinking the coefficients values to zero and thus helps in feature selection procedure
  - Also called L1 regularization
    - Regularization: helps in reducing the complexity of the model
    - L1: Absolute values of the coefficients

- L1 regularization term: $\lambda *(|\beta_0| + |\beta_1| + \cdots + |\beta_p|)$
  - λ : penalty term

LASSO: Objective Function

- Find the values of the coefficients that minimize the *sum of the squared differences between the predicted values and the actual values* and *L1 regularization term*
  - Minimize: RSS+ λ *($|\beta_0| + |\beta_1| + \cdots + |\beta_p|$)



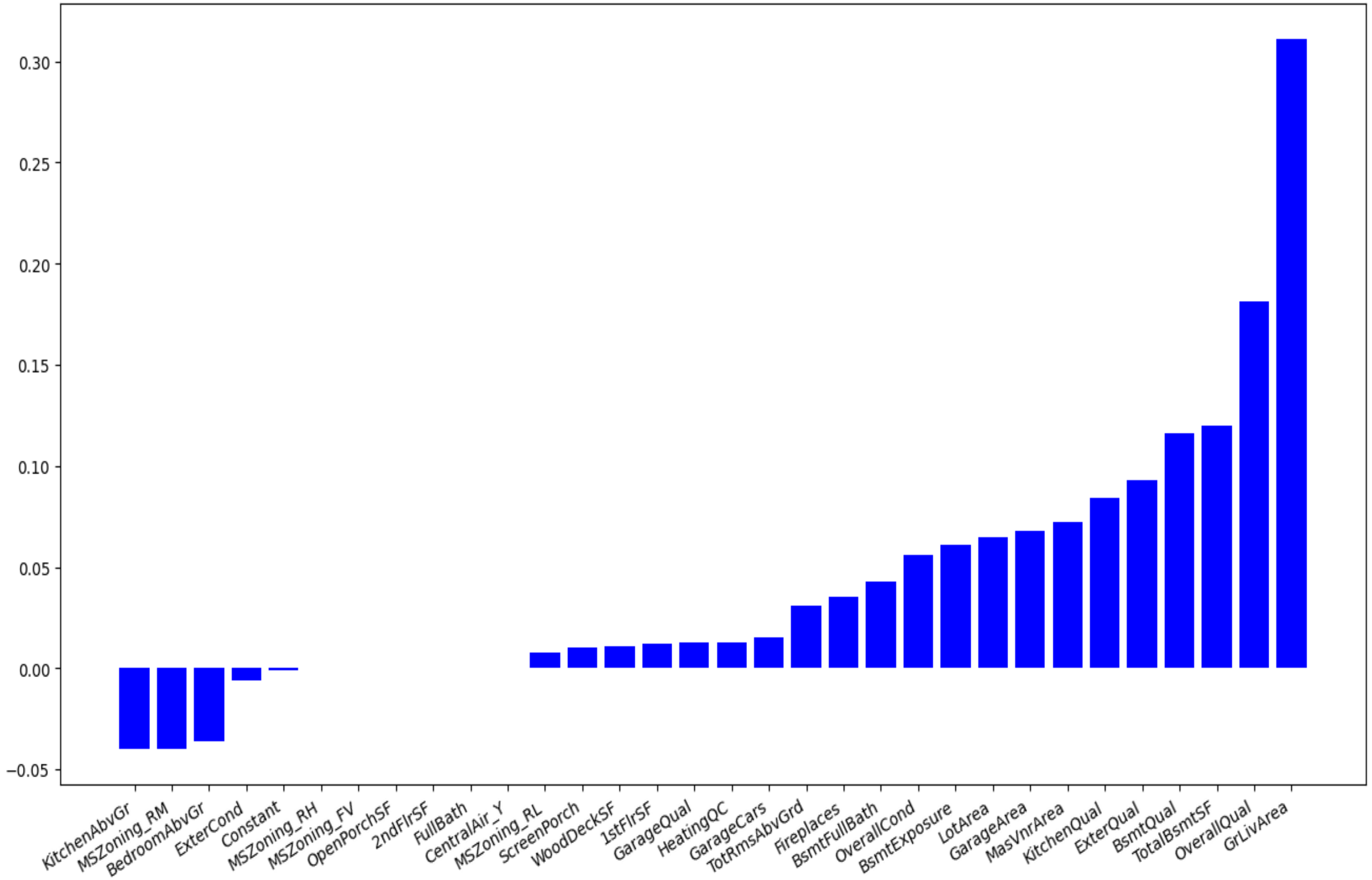
LASSO regression estimates are given by the first point at which an ellipse contacts the diamond region

LASSO: Shrinkage Property

- LASSO shrinks coefficient estimates to zero
  - Useful for feature selection, as the variables with zero coefficients are effectively removed from the model.

- The choice of the regularization parameter λ is crucial in LASSO regression.
  - A larger λ value increases the amount of regularization, leading to more coefficients being pushed towards zero.
  - A smaller λ value reduces the regularization effect, allowing more variables to have non-zero coefficients.
  - λ can be obtained through cross-validation technique

LASSO- Parameter Results

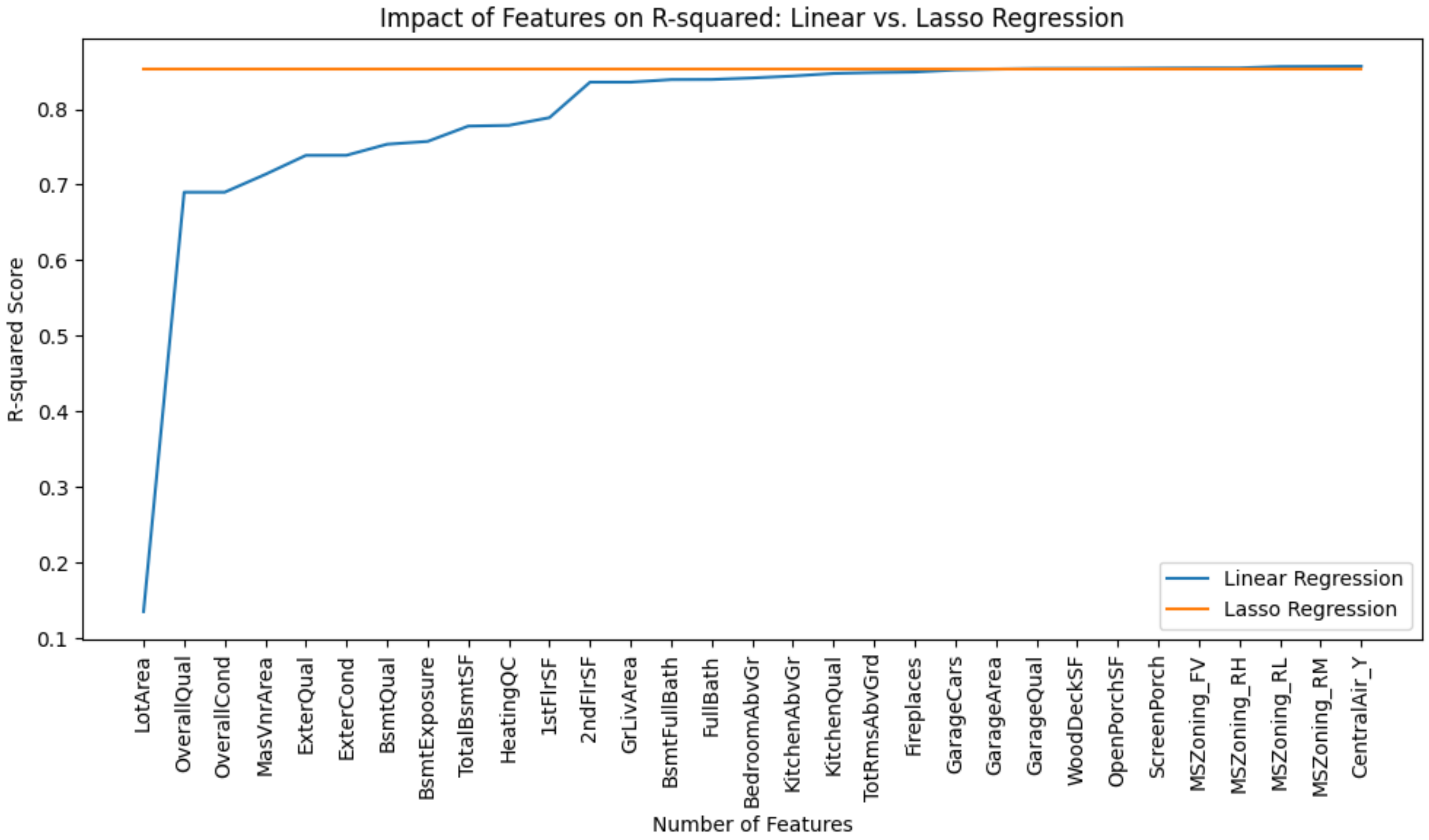



```python
from sklearn.linear_model import Lasso

lm = Lasso(alpha=0.01)
lm.fit(X_train,y_train)
y = lm.predict(X)
```

Feature Selection in Linear Regression

- Simple Way: Adding Features iteratively to identify the best sub-set



Thank You

© Copyright National University of Singapore. All Rights Reserved.