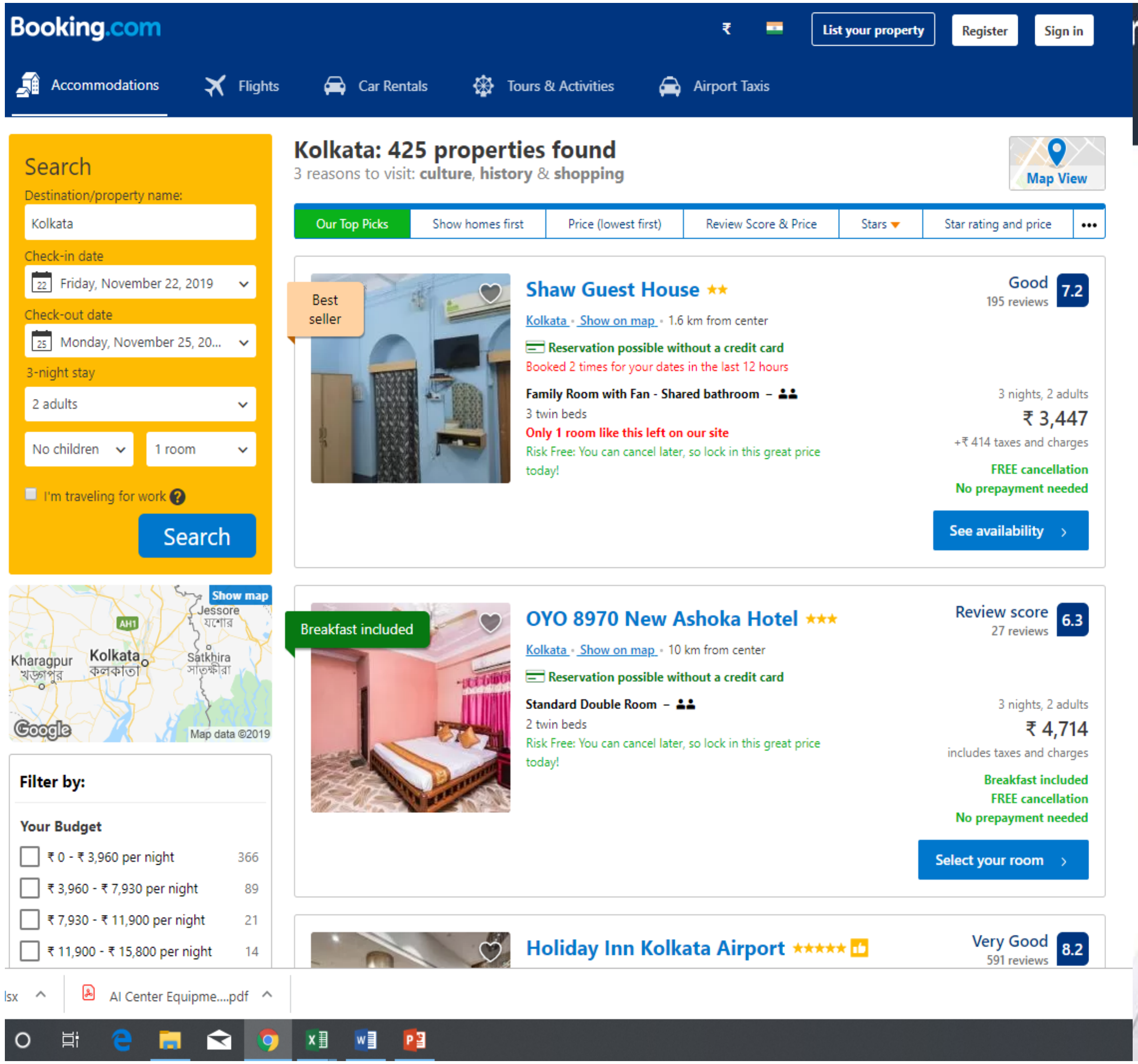
Regularized Linear Regression and Decision Trees

Artificial Intelligence for Economics (AI60003)

Module 2, Lecture 4

Adway Mitra

Average Ratings



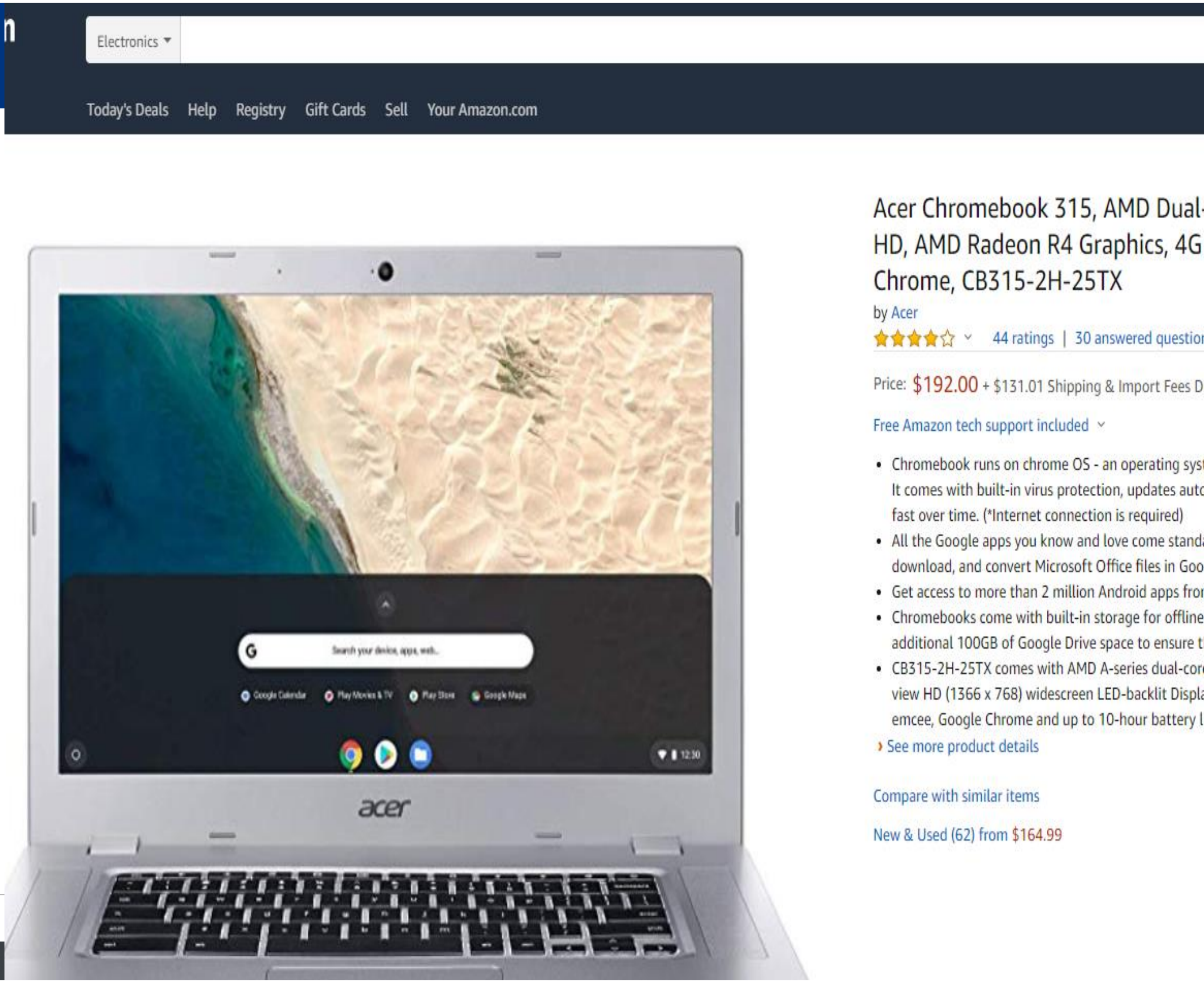
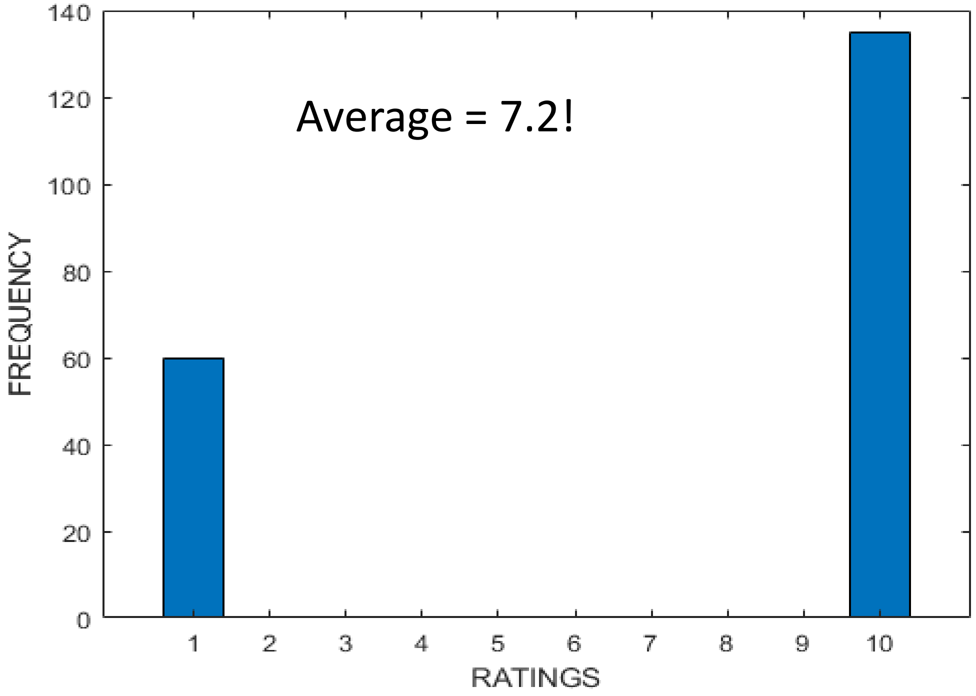
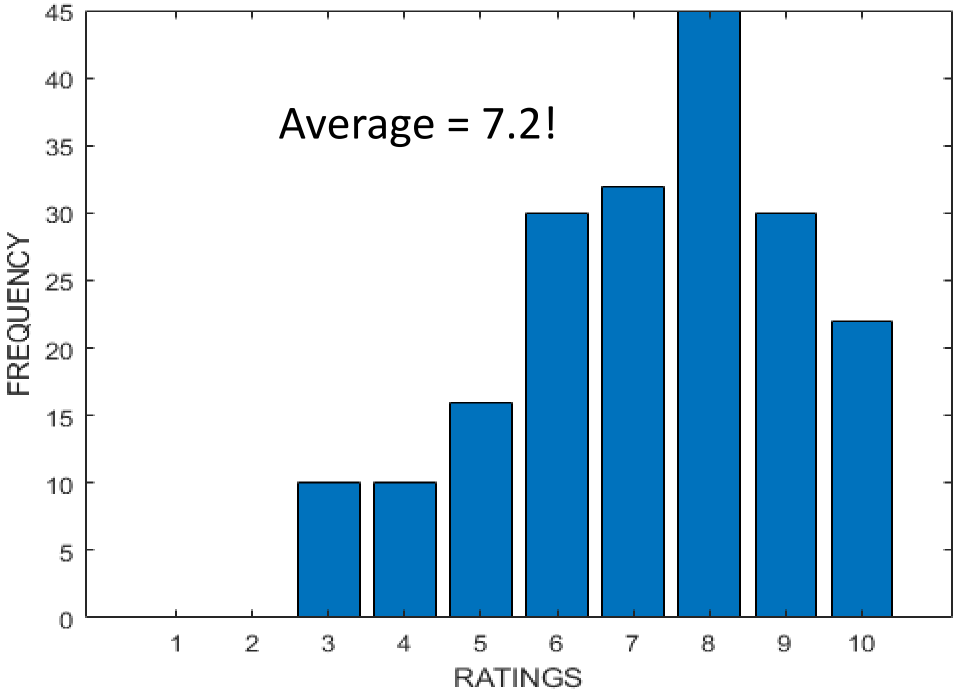


Image source: Google Images

Average Ratings

- 195 reviews, on a scale of 1 to 10
- Average rating: 7.2!
- There may be large or small variance among individual reviews





How do users rate a product?

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor
1. The relevance of this topic to me was	4	3	2	1
2. The usefulness of materials was	4	3	2	1
3. The effectiveness of the presenter was	4	3	2	1
4. I expect the future usefulness of this topic to be	4	3	2	1
5. My overall evaluation of this session is	4	3	2	1

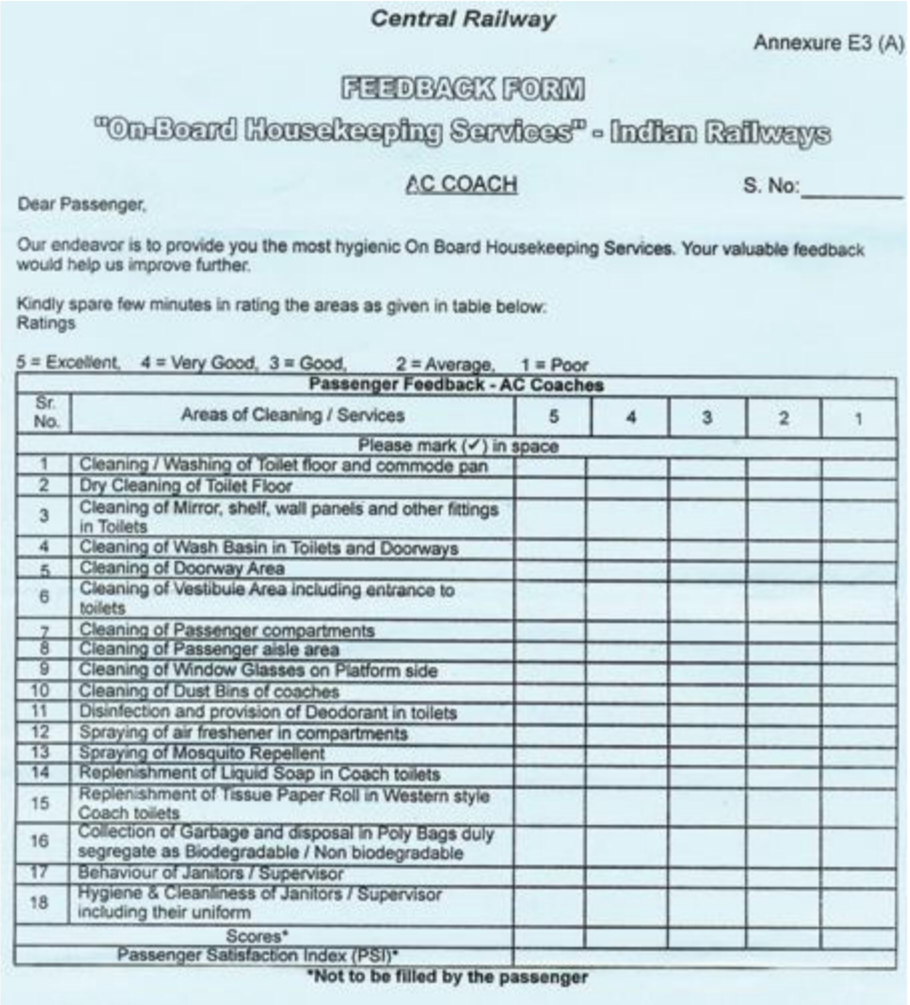
Your Account > Packaging Feedback

Rate Amazon's Packaging

Did the packaging protect your items adequately?	☆☆☆☆☆ Protection	1 star = Poor; 5 stars = Excellent
Was the box size and packaging appropriate for the items?	○ Too Small ○ About Right ○ Too Big ○ Way Too Big	

Rate Item's Packaging

	☆☆☆☆☆ Ease of Opening	1 star = Very Difficult; 5 stars = Very Easy



Central Railway

Annexure E3 (A)

FEEDBACK FORM

"On-Board Housekeeping Services" - Indian Railways

AC COACH

S. No:________

Dear Passenger,

Our endeavor is to provide you the most hygienic On Board Housekeeping Services. Your valuable feedback would help us improve further.

Kindly spare few minutes in rating the areas as given in table below:
Ratings

5 = Excellent, 4 = Very Good, 3 = Good, 2 = Average, 1 = Poor

Passenger Feedback - AC Coaches

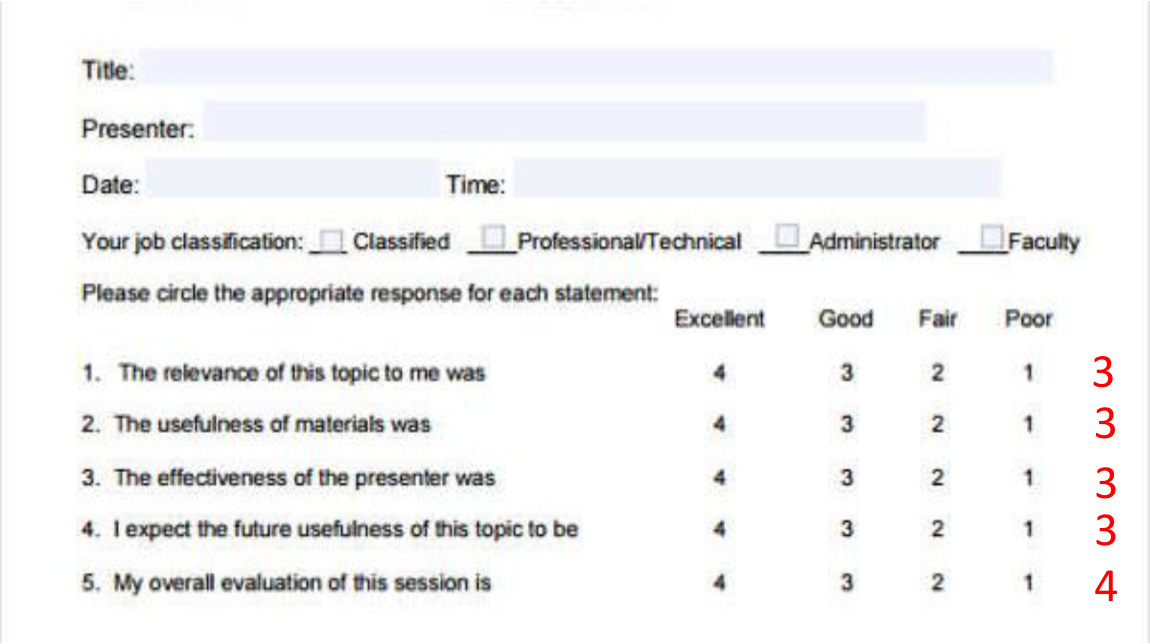
Sr. No.	Areas of Cleaning / Services	5	4	3	2	1
	Please mark (✓) in space					
1	Cleaning / Washing of Toilet floor and commode pan					
2	Dry Cleaning of Toilet Floor					
3	Cleaning of Mirror, shelf, wall panels and other fittings in Toilets					
4	Cleaning of Wash Basin in Toilets and Doorways					
5	Cleaning of Doorway Area					
6	Cleaning of Vestibule Area including entrance to toilets					
7	Cleaning of Passenger compartments					
8	Cleaning of Passenger aisle area					
9	Cleaning of Window Glasses on Platform side					
10	Cleaning of Dust Bins of coaches					
11	Disinfection and provision of Deodorant in toilets					
12	Spraying of air freshener in compartments					
13	Spraying of Mosquito Repellent					
14	Replenishment of Liquid Soap in Coach toilets					
15	Replenishment of Tissue Paper Roll in Western style Coach toilets					
16	Collection of Garbage and disposal in Poly Bags duly segregate as Biodegradable / Non biodegradable					
17	Behaviour of Janitors / Supervisor					
18	Hygiene & Cleanliness of Janitors / Supervisor including their uniform					
	Scores*					
	Passenger Satisfaction Index (PSI)*					

***Not to be filled by the passenger**

Image source: Google Images

How do users rate a product?

User 1:



Title:

Presenter:

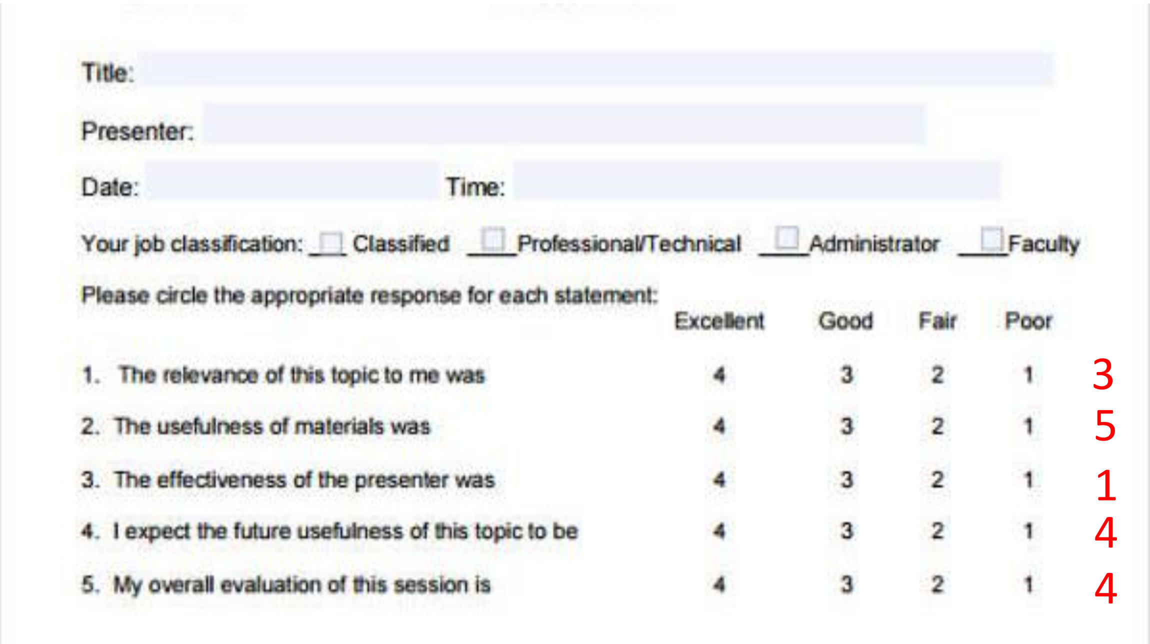
Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	3
2. The usefulness of materials was	4	3	2	1	3
3. The effectiveness of the presenter was	4	3	2	1	3
4. I expect the future usefulness of this topic to be	4	3	2	1	3
5. My overall evaluation of this session is	4	3	2	1	4

User 2:



Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	3
2. The usefulness of materials was	4	3	2	1	5
3. The effectiveness of the presenter was	4	3	2	1	1
4. I expect the future usefulness of this topic to be	4	3	2	1	4
5. My overall evaluation of this session is	4	3	2	1	4

Image source: Google Images

How do users rate a product?

- Each product has N features ($f_1, f_2, \ldots, f_N$)
- The rating "y_i" given by any user "i" may be a weighted average of her scores ($x_{i1}, x_{i2}, \ldots, x_{iN}$) on the individual features
- The weights ($w_{i1}, w_{i2}, \ldots, w_{iN}$) may vary from one user to another according to their respective priorities
- Simplest model for user rating: $y_i = \sum_j w_{ij}x_{ij} + b_i$ (b_i: bias)

How do users rate a product?

- Each product has N features ($f_1, f_2, \ldots, f_N$)
- The rating "y_i" given by any user "i" may be a weighted average of her scores ($x_{i1}, x_{i2}, \ldots, x_{iN}$) on the individual features
- The weights ($w_{i1}, w_{i2}, \ldots, w_{iN}$) may vary from one user to another according to their respective priorities
- Simplest model for user rating: $y_i = \sum_j w_{ij}x_{ij} + b_i$ (b_i: bias)
- Need to estimate the weights "w": M users x N features
- Too many parameters!!

How do users rate a product?

- Each product has N features (f_1, f_2, …. , f_N)
- The rating "y_i" given by any user "i" may be a weighted average of her scores (x_{i1}, x_{i2}, … , x_{iN}) on the individual features
- The weights (w_{i1}, w_{i2}, … , w_{iN}) may vary from one user to another according to their respective priorities
- Simplest model for user rating: $y_i = \sum_j w_{ij}x_{ij} + b_i$ (b_i: bias)
- Need to estimate the weights "w": M users x N features
- Too many parameters!!
- New approximate model: $y_i = \sum_j w_j x_{ij} + b$, i.e. all users have equal weights!

Linear Regression

- We know the feature scores "s_{ij}" and the final score "x_i"
- We want to find out the relative importance of the different features (on average)
- The answer: linear regression!
- General Recipe:

1) Define a model with parameters (w, b)
2) Define a measure on how well the model can fit the final scores
3) Choose the model parameters to improve this measure!

Linear Regression

- The model in this case: $h_i = \sum_j w_j x_{ij} + b$ (h_i: predicted rating)
- Measurement of fit: squared error loss function
- $L(y_i, h_i) = (y_i - h_i)^2 = \sum_j (y_i - w_j x_{ij} - b)^2$

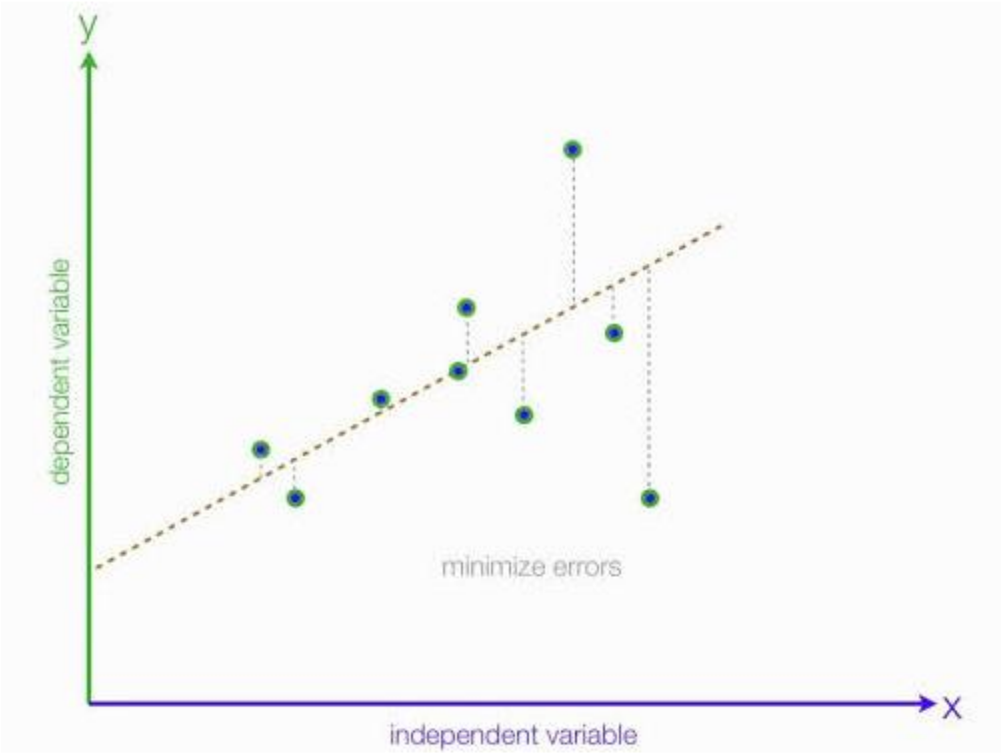


Image source: Google Images

Linear Regression

- The model in this case: $h_i = \sum_j w_j x_{ij} + b$ (h_i: predicted rating)
- Measurement of fit: squared error loss function
- Loss for user i: $L(y_i, h_i) = (y_i - h_i)^2 = \sum_j (y_i - w_j x_{ij} - b)^2$

- Choose w, b to minimize total loss $\sum_i L(y_i, h_i)$ over all M users!
- Differentiate the total loss w.r.t. each variable, equate to 0, and solve an equation!

Linear Regression in one dimension

First, let us consider each product has only one feature

$\frac{dL}{dw} = 0 \implies 2\sum_i (y_i - wx_i - b)x_i = 0$

$\frac{dL}{db} = 0 \implies 2\sum_i (y_i - wx_i - b) = 0$

Solving these equations, we get

$b = \bar{y} - w\bar{x}$

$w = (\sum_i (\tilde{x}_i)^2)^{-1} (\sum_i \tilde{x}_i \tilde{y}_i)$

where $\bar{x} = \frac{1}{N}\sum_i x_i$, $\bar{y} = \frac{1}{N}\sum_i y_i$, $\tilde{x}_i = x_i - \bar{x}$

In [3]:

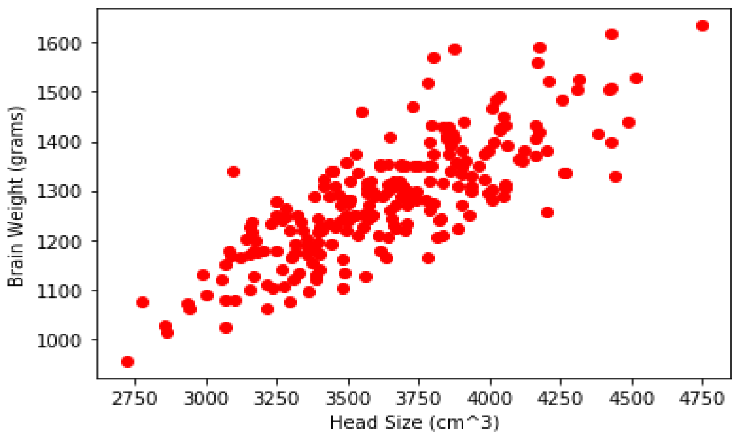
```
#initializing our inputs and outputs


#mean of our inputs and outputs
x_mean = np.mean(X)
y_mean = np.mean(Y)

#total number of values
n = len(X)

#using the formula to calculate the b1 and b0
numerator = 0
denominator = 0
for i in range(n):
    numerator += (X[i] - x_mean) * (Y[i] - y_mean)
    denominator += (X[i] - x_mean) ** 2

b1 = numerator / denominator
b0 = y_mean - (b1 * x_mean)

#printing the coefficient
print(b1, b0)
```

Python Implementation

In [2]:
```
#import libraries
%matplotlib inline
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

#reading data
dataset = pd.read_csv('dataset.csv')
print(dataset.shape)
dataset.head()

X = dataset['Head Size(cm^3)'].values
Y = dataset['Brain Weight(grams)'].values

#plot the data point
plt.scatter(X, Y, color='#ff0000', label='Data Point')

# x-axis label
plt.xlabel('Head Size (cm^3)')

#y-axis label
plt.ylabel('Brain Weight (grams)')
```

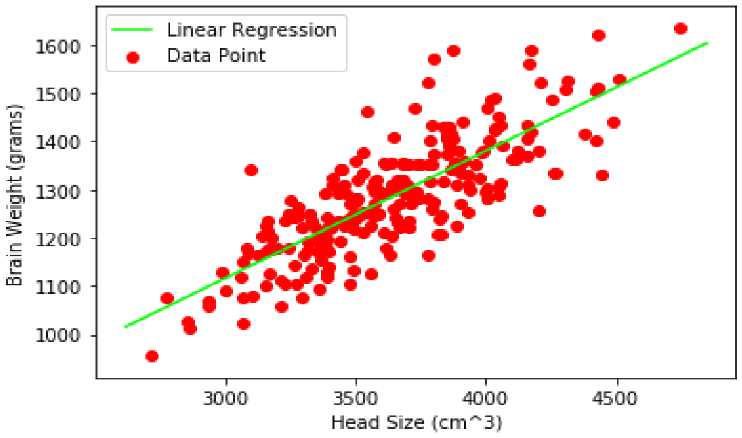
```
(237, 4)
```

Out[2]: `Text(0, 0.5, 'Brain Weight (grams)')`




```
#mean of our inputs and outputs
x_mean = np.mean(X)
y_mean = np.mean(Y)

#total number of values
n = len(X)

#using the formula to calculate the b1 and b0
numerator = 0
denominator = 0
for i in range(n):
    numerator += (X[i] - x_mean) * (Y[i] - y_mean)
    denominator += (X[i] - x_mean) ** 2

b1 = numerator / denominator
b0 = y_mean - (b1 * x_mean)

#printing the coefficient
print(b1, b0)
```

In [3]:
```
#plotting values
x_max = np.max(X) + 100
x_min = np.min(X) - 100

#calculating line values of x and y
x = np.linspace(x_min, x_max, 1000)
y = b0 + b1 * x

plt.plot(x, y, color='#00ff00', label='Linear Regression') #plotting line
plt.scatter(X, Y, color='#ff0000', label='Data Point') #plot the data point
plt.xlabel('Head Size (cm^3)') # x-axis label
plt.ylabel('Brain Weight (grams)') #y-axis label

plt.legend()
plt.show()
```



Average Rating Prediction

- Given a new product, we need to predict it's "average rating"
- Average rating = $mean_i(y_i)$
- According to LR model:
- predicted average rating = $mean_i(h_i)$
- $= mean_i(\sum_j w_j x_{ij} + b) = \sum_j w_j \, mean_i(x_{ij}) + b$
- We have the weights "w_j" of its features and bias "b", by linear regression for <u>similar products</u>
- We can find the average user ratings of each feature $mean_i(x_{ij})$, based on <u>other products having same feature</u>

Average Rating Prediction

- New Product: a new camera model
- Features: resolution, battery life, memory, flash, weight, size
- Weights of features: calculate by linear regression from user ratings on other cameras
- New camera resolution: 5 MP
- Average rating on resolution: 4.0
- Weight of resolution: 0.54

Model	Resolution	Mean feature rating
Camera1	5 MP	4.1
Camera2	5 MP	3.9
Camera3	10 MP	4.4
Camera4	12 MP	4.1
Camera5	6 MP	4.0
Camera6	15 MP	4.3

Average Rating Prediction

- New Product: a new camera model
- Features: resolution, battery life, memory, flash, weight, size
- Weights of features: calculate by linear regression from user ratings on other cameras
- New camera battery life: 2 years
- Average rating on battery life : 3.8
- Weight of battery life : 0.36

Model	Battery Life	Mean feature rating
Camera1	3 years	4.5
Camera2	2 years	3.6
Camera3	2 years	3.8
Camera4	1 year	3.1
Camera5	2 years	3.9
Camera6	3 years	4.3

Average Rating Prediction

- New Product: a new camera model
- Features: resolution, battery life, memory, flash, weight, size
- Weights of features: calculate by linear regression from user ratings on other cameras
- New camera memory: 5 GB
- Average rating on memory: 4.5
- Weight of memory: 0.10

- Predicted average rating

= 0.54*4.0 + 0.36*3.8 + 0.1*4.5 = 4.0!

Model	Memory	Mean feature rating
Camera1	1 GB	3.8
Camera2	1 GB	3.9
Camera3	2 GB	4.1
Camera4	3 GB	4.0
Camera5	5 GB	4.4
Camera6	5 GB	4.5

How do users rate a product?

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor
1. The relevance of this topic to me was	4	3	2	1
2. The usefulness of materials was	4	3	2	1
3. The effectiveness of the presenter was	4	3	2	1
4. I expect the future usefulness of this topic to be	4	3	2	1
5. My overall evaluation of this session is	4	3	2	1

Your Account > Packaging Feedback

Rate Amazon's Packaging

Did the packaging protect your items adequately?	☆☆☆☆☆ Protection	1 star = Poor; 5 stars = Excellent
Was the box size and packaging appropriate for the items?	○ Too Small ○ About Right ○ Too Big ○ Way Too Big	

Rate Item's Packaging

	☆☆☆☆☆ Ease of Opening	1 star = Very Difficult; 5 stars = Very Easy

Central Railway

Annexure E3 (A)

FEEDBACK FORM

"On-Board Housekeeping Services" - Indian Railways

AC COACH S. No:__________

Dear Passenger,

Our endeavor is to provide you the most hygienic On Board Housekeeping Services. Your valuable feedback would help us improve further.

Kindly spare few minutes in rating the areas as given in table below:
Ratings

5 = Excellent, 4 = Very Good, 3 = Good, 2 = Average, 1 = Poor

Passenger Feedback - AC Coaches

Sr. No.	Areas of Cleaning / Services	5	4	3	2	1
	Please mark (✓) in space					
1	Cleaning / Washing of Toilet floor and commode pan					
2	Dry Cleaning of Toilet Floor					
3	Cleaning of Mirror, shelf, wall panels and other fittings in Toilets					
4	Cleaning of Wash Basin in Toilets and Doorways					
5	Cleaning of Doorway Area					
6	Cleaning of Vestibule Area including entrance to toilets					
7	Cleaning of Passenger compartments					
8	Cleaning of Passenger aisle area					
9	Cleaning of Window Glasses on Platform side					
10	Cleaning of Dust Bins of coaches					
11	Disinfection and provision of Deodorant in toilets					
12	Spraying of air freshener in compartments					
13	Spraying of Mosquito Repellent					
14	Replenishment of Liquid Soap in Coach toilets					
15	Replenishment of Tissue Paper Roll in Western style Coach toilets					
16	Collection of Garbage and disposal in Poly Bags duly segregate as Biodegradable / Non biodegradable					
17	Behaviour of Janitors / Supervisor					
18	Hygiene & Cleanliness of Janitors / Supervisor including their uniform					
	Scores*					
	Passenger Satisfaction Index (PSI)*					

*Not to be filled by the passenger

Image source: Google Images

How do users rate a product?

User 1:

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	2
2. The usefulness of materials was	4	3	2	1	2
3. The effectiveness of the presenter was	4	3	2	1	5
4. I expect the future usefulness of this topic to be	4	3	2	1	2
5. My overall evaluation of this session is	4	3	2	1	4

User 2:

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	4
2. The usefulness of materials was	4	3	2	1	4
3. The effectiveness of the presenter was	4	3	2	1	1
4. I expect the future usefulness of this topic to be	4	3	2	1	3
5. My overall evaluation of this session is	4	3	2	1	2

For both users, feature 3 seems to play a major role in deciding the overall evaluation, other features have smaller impact

Image source: Google Images

How do users rate a product?

User 1:

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	1
2. The usefulness of materials was	4	3	2	1	2
3. The effectiveness of the presenter was	4	3	2	1	5
4. I expect the future usefulness of this topic to be	4	3	2	1	1
5. My overall evaluation of this session is	4	3	2	1	5

User 2:

Title:

Presenter:

Date: Time:

Your job classification: ☐ Classified ☐ Professional/Technical ☐ Administrator ☐ Faculty

Please circle the appropriate response for each statement:

	Excellent	Good	Fair	Poor	
1. The relevance of this topic to me was	4	3	2	1	5
2. The usefulness of materials was	4	3	2	1	4
3. The effectiveness of the presenter was	4	3	2	1	1
4. I expect the future usefulness of this topic to be	4	3	2	1	5
5. My overall evaluation of this session is	4	3	2	1	1

For both users, feature 3 seems to be the only factor in deciding the overall evaluation, other features do not matter

Image source: Google Images

Feature Selection

- Linear regression model: $y_i = \sum_j w_j x_{ij} + b_i$, i.e. all feature ratings contribute to the final rating
- But in the examples, only a small number of features seem to influence the final rating, other features have little importance
- In case 1: One element in "w" will have high value, other elements will have small values
- In case 2: All elements except one in "w" have 0 value, i.e. "w" is sparse!

Feature Selection

- Feature selection: the task of identifying the "important" features
- Important feature: those which strongly influence the final ratings
- In the given examples, feature selection is easy by manual inspection
- Large dataset: many examples, many dimensions, noisy ratings, manual inspection impossible
- Can linear regression itself solve the feature selection problem?
- It can, if it returns a suitable "w"!

Sparse Regression for Feature Selection

- Case 1: we want “w” such that most of its elements are small
- Case 2: we want “w” such that most of its elements are 0
- Can we convert these demands into mathematical formulations?

Sparse Regression for Feature Selection

- Case 1: we want "w" such that most of its elements are small
- Case 2: we want "w" such that most of its elements are 0
- Can we convert these demands into mathematical formulations?
- General recipe: find a regularization function f(w)
- f(w) should have low value for suitable "w", high value for unsuitable "w"

Sparse Regression for Feature Selection

- Case 1: we want “w” such that most of its elements are small
- Case 2: we want “w” such that most of its elements are 0
- Can we convert these demands into mathematical formulations?
- General recipe: find a regularization function $f(w)$
- $f(w)$ should have low value for suitable “w”, high value for unsuitable “w”

- Find (w,b) to minimize $L(w,b) + \lambda f(w)$
- First term to find w that fits data, second term to find “w” that is suitable, λ to balance them!

LASSO regression

- Our original aim: “sparse w”!
- The L_0-norm of vector “w”: number of non-zero elements
- Regularizer $f(w) = ||w||_0$ promotes sparse “w”!
- New problem: $L(w,b) + \lambda f(w)$
- Non-differentiable function!!!

LASSO regression

- Our original aim: “sparse w”!
- The L_0-norm of vector “w”: number of non-zero elements
- Regularizer $f(w) = ||w||_0$ promotes sparse “w”!
- New problem: $L(w,b) + \lambda f(w)$
- Non-continuous function!!!
- Relaxation: $f(w) = ||w||_1 = \sum_j |w_j|$ = sum of absolute values of elements!
- Low value of $||w||_1$: most values of w “close to 0”
- “Almost sparse” w!

LASSO vs Ridge Regression

- Both are compromise between squared loss minimization and feasible region

  Feasible region shape different in both cases

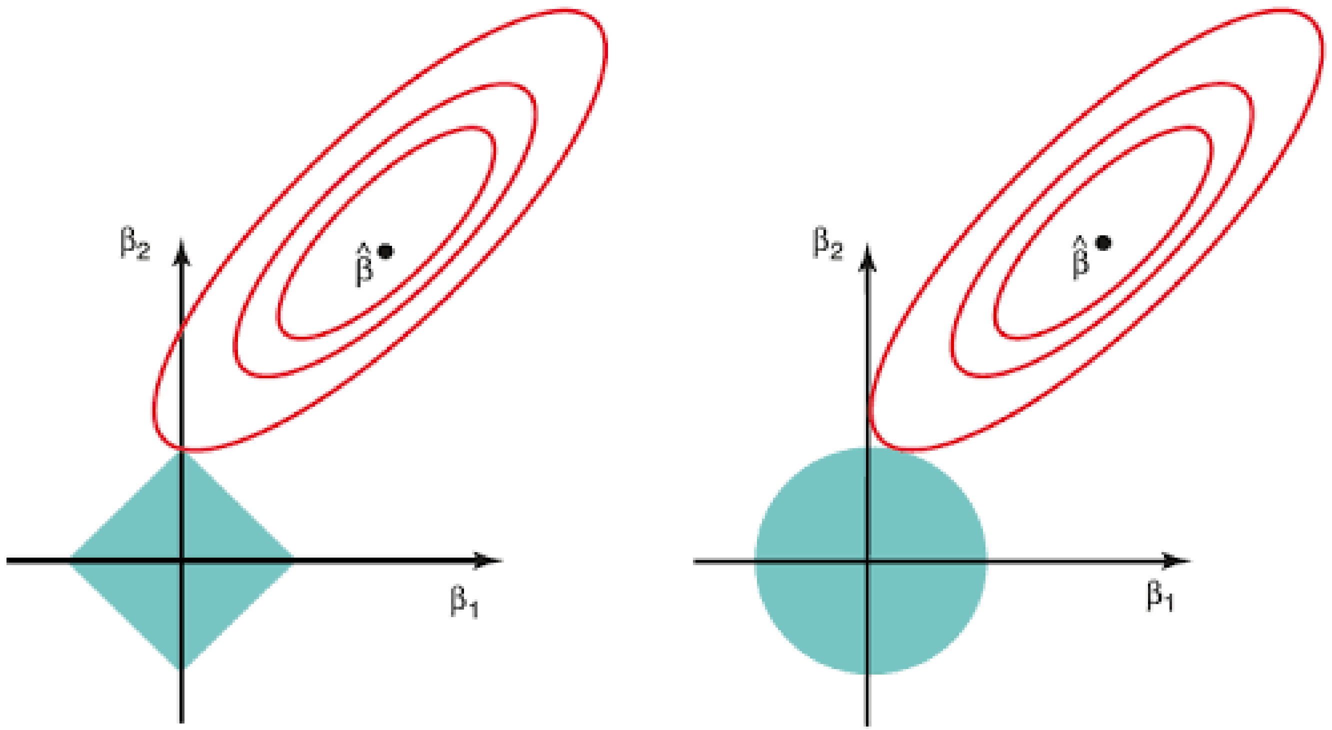


Image source: Google Images

LASSO regression

- Objective function: $\sum_i (y_i - w^T x_i - b)^2 + \lambda ||w||_1$
- Difficult to solve by differentiation!

- Alternative: use numerical method instead of analytical!
- Gradient Descent: to be covered later!

Python Implementation using sklearn

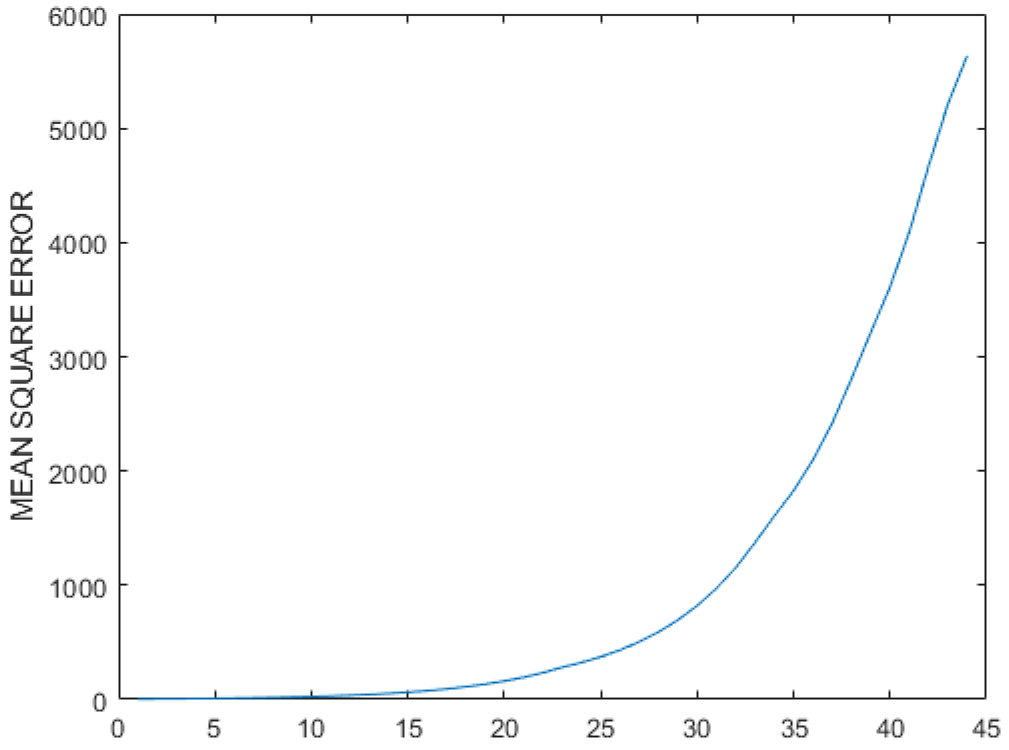
In [64]:
```
TrainX=np.asarray(X)
TrainY=np.asarray(Y)

type(NewX)
```

Out[64]: numpy.ndarray

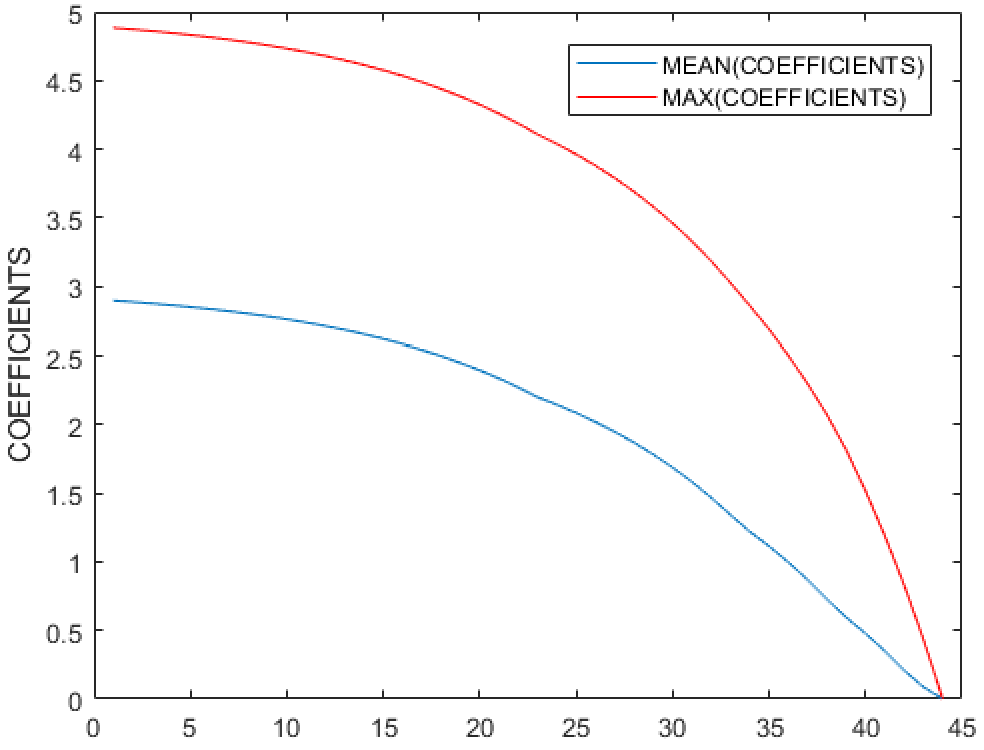
In [0]:
```
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model  import Lasso
from sklearn.linear_model  import Ridge
```

In [73]:
```
lasso=Lasso()
parameters={'alpha': [0.001,0.01,0.1, 0.5,1]}
lassoReg=GridSearchCV(lasso,parameters,scoring='neg_mean_squared_error',cv=3)    #using gridsearch for cross validation
lassoReg.fit(TrainX.reshape(-1,1),TrainY.reshape(-1,1))    # training

ridge=Ridge()
parameters={'alpha': [0.1, 0.5,1]}
ridgeReg=GridSearchCV(ridge,parameters,scoring='neg_mean_squared_error',cv=3)    #using gridsearch for cross validation
ridgeReg.fit(TrainX.reshape(-1,1),TrainY.reshape(-1,1))    # training
```

LASSO regression

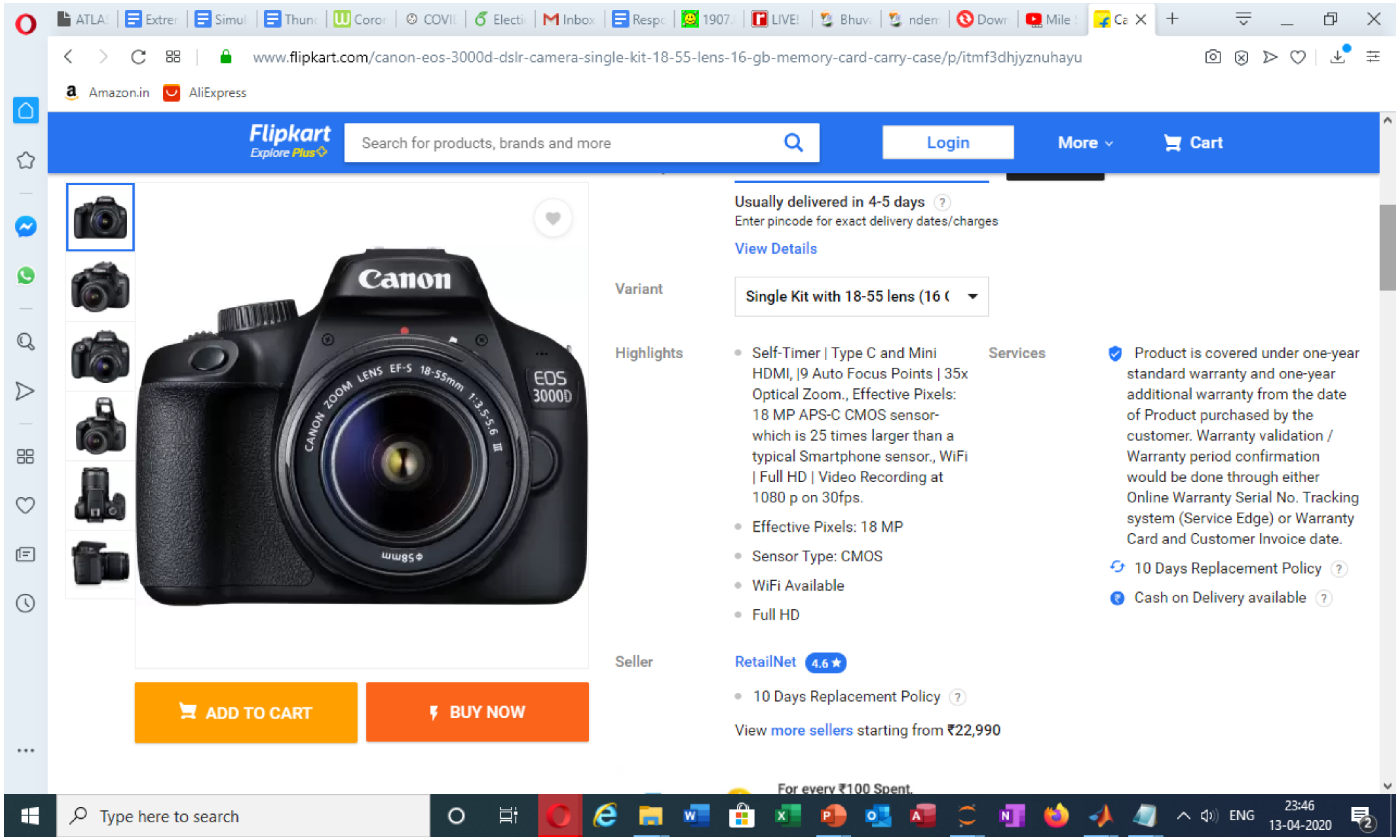


LAMBDA



LAMBDA

Discrete Product Ratings based on Discrete Features



How much rating will a particular user give this camera out of 5?

Probably depends on features!

Which features does the user like?

Source: Flipkart website

Feature Selection

- The user has exactly 5 options: 1, 2, 3, 4 or 5 stars!
- Her choice depends on the different features of the product!
- But she may consider some features to be more important than others !
- Which features determine her vote?

Company	Color	Resolution	Video Rate	Price	Her Rating
C1	Black	10 MP	25 fps	$200	2
C1	White	15 MP	25 fps	$250	2
C2	White	12 MP	30 fps	$250	4
C1	Black	15 MP	30 fps	$300	3
C2	Black	20 MP	25 fps	$400	3
C2	White	12 MP	50 fps	$500	5
C2	Black	15 MP	30 fps	$250	????

Feature Selection

- The user has 5 exactly options: 1, 2, 3, 4 or 5 stars!
- Her choice depends on the different features of the product!
- But she may consider some features to be more important than others !
- Which features determine her vote?

Company	Color	Resolution	Video Rate	Price	Her Rating
C1	Black	10 MP	25 fps	$200	2
C1	White	15 MP	25 fps	$250	2
C2	White	12 MP	30 fps	$250	4
C1	Black	15 MP	30 fps	$300	3
C2	Black	20 MP	25 fps	$400	3
C2	White	12 MP	50 fps	$500	5
C2	Black	15 MP	30 fps	$350	4

Feature Selection

- The user has 5 exactly options: 1, 2, 3, 4 or 5 stars!
- Her choice depends on the different features of the product!
- But she may consider some features to be more important than others !
- Which features determine her vote?

Company	Color	Resolution	Video Rate	Price	Her Rating
C1	Black	10 MP	25 fps	$200	2
C1	White	15 MP	25 fps	$250	2
C2	White	12 MP	30 fps	$250	4
C1	Black	15 MP	30 fps	$300	3
C2	Black	20 MP	25 fps	$400	3
C2	White	12 MP	50 fps	$500	5
C2	Black	15 MP	30 fps	$350	4

Decision Tree for Feature Selection

- Which features does she consider as important while rating?
- Let's look at her history of rating 100 cameras!

Rating	Count
1	21
2	24
3	18
4	20
5	17

Overall,
Count=100

Rating	Count
1	15
2	18
3	10
4	5
5	6

Company = C1,
Count=54

Rating	Count
1	6
2	6
3	8
4	15
5	11

Company = C2,
Count=46

Rating	Count
1	15
2	20
3	13
4	12
5	10

Color=Black,
Count=70

Rating	Count
1	6
2	4
3	5
4	8
5	7

Color=White,
Count=30

Decision Tree for Feature Selection

- Which features does she consider as important while rating?
- Let's look at her history of rating 100 cameras!

Rating	Count
1	21
2	24
3	18
4	20
5	17

Overall,
Count=100

Rating	Count
1	15
2	18
3	10
4	5
5	6

Company = C1,
Count=54

Rating	Count
1	6
2	6
3	8
4	15
5	11

Company = C2,
Count=46

Rating	Count
1	15
2	20
3	13
4	12
5	10

Color=Black,
Count=70

Rating	Count
1	6
2	4
3	5
4	8
5	7

Color=White,
Count=30

Which feature is more important for ratings - company or color???

What’s a discriminative feature?

- Company ={C1, C2}, Price = real number, Y = {LOW (1-3), HIGH (4-5)}

	COMPANY=C1	COMPANY=C2	
#(Y=LOW)	43	20	63
#(Y=HIGH)	11	26	37
Total	54	46	100

What’s a discriminative feature?

- Company ={C1, C2}, Price = real number, Y = {LOW (1-3), HIGH (4-5)}

	Price<300	Price >=300	
#(Y=LOW)	45	18	63
#(Y=HIGH)	25	12	37
Total	70	30	100

What’s a discriminative feature?

- Company ={C1, C2}, Price = real number, Y = {LOW (1-3), HIGH (4-5)}

	Price<500	**Price >=500**	
#(Y=LOW)	55	8	63
#(Y=HIGH)	35	2	37
Total	90	10	100

What’s a discriminative feature?

- Prob(Y = HIGH | COMPANY = C1) = 11/54 ~ 0.2 [Easy to decide]
- Prob(Y = HIGH | COMPANY = C2) = 26/46 ~ 0.55

- Prob(Y = HIGH | PRICE < 300) = 25/70 ~ 0.36
- Prob(Y = HIGH | PRICE >= 300) = 12/30 = 0.4

- Prob(Y = HIGH | PRICE < 500) = 35/90 ~ 0.4
- Prob(Y = HIGH | PRICE >= 500) = 2/10 ~ 0.2 [Easy to decide][Very few examples]

What’s a discriminative feature?

- Prob(Y = HIGH | COMPANY = C1) = 11/54 ~ 0.2 [Easy to decide]
- Prob(Y = HIGH | COMPANY = C1) = 26/46 ~ 0.55

COMPANY: good feature

- Prob(Y = HIGH | PRICE < 300) = 25/70 ~ 0.36
- Prob(Y = HIGH | PRICE >= 300) = 12/30 = 0.4

PRICE<300: bad feature

- Prob(Y = HIGH | PRICE < 500) = 35/90 ~ 0.4
- Prob(Y = HIGH | PRICE >= 500) = 2/10 ~ 0.2 [Easy to decide][Very few examples]

PRICE<500: doubtful feature

Decision Tree Algorithm

- Idea: identify the “most discriminative” feature, use it to classify!
- Problem 1: How to quantify “discriminative-ness”?
- Problem 2: What if no feature is very discriminative?

Decision Tree Algorithm

- Idea: identify the “most discriminative” feature, use it to classify!

- Problem 1: How to quantify “discriminative-ness”?

  - entropy!

- Problem 2: What if no feature is very discriminative?

  - try a sequence of features!

Entropy: measure of discriminativeness

- P(Y=1) = 0.5, p(Y=2) = 0.5 : low discriminative ability
- P(Y=1) = 0.9, p(Y=2) = 0.1 : high discriminative ability

$$\boxed{H = -\sum_i p_i(\log_2 p_i)}$$

- Case 1: H = 1 (0.69)
- Case 2: H = 0.5 (0.33)

Feature selection based on entropy

	COMPANY=C1	COMPANY=C2	NO SPLIT
#(Y=LOW)	43	20	63
#(Y=HIGH)	11	26	37
Entropy	0.51	0.68	0.66

- Information gain =

Original Entropy – (Split1_size*Split1_ entropy + Split2_size*Split2_ entropy)

0.66 – (54/100*0.51 + 46/100*0.68) ~ 0.07

Feature selection based on entropy

	PRICE<300	PRICE>=300	NO SPLIT
#(Y=LOW)	45	18	63
#(Y=HIGH)	25	12	37
Entropy	0.65	0.67	0.66

- Information gain =

Original Entropy – (Split1_size*Split1_ entropy + Split2_size*Split2_ entropy)

0.66 – (70/100*0.65 + 30/100*0.67) ~ 0!!

Feature selection based on entropy

	PRICE<500	PRICE>=500	NO SPLIT
#(Y=LOW)	55	8	63
#(Y=HIGH)	35	2	37
Entropy	0.67	0.5	0.66

- Information gain =

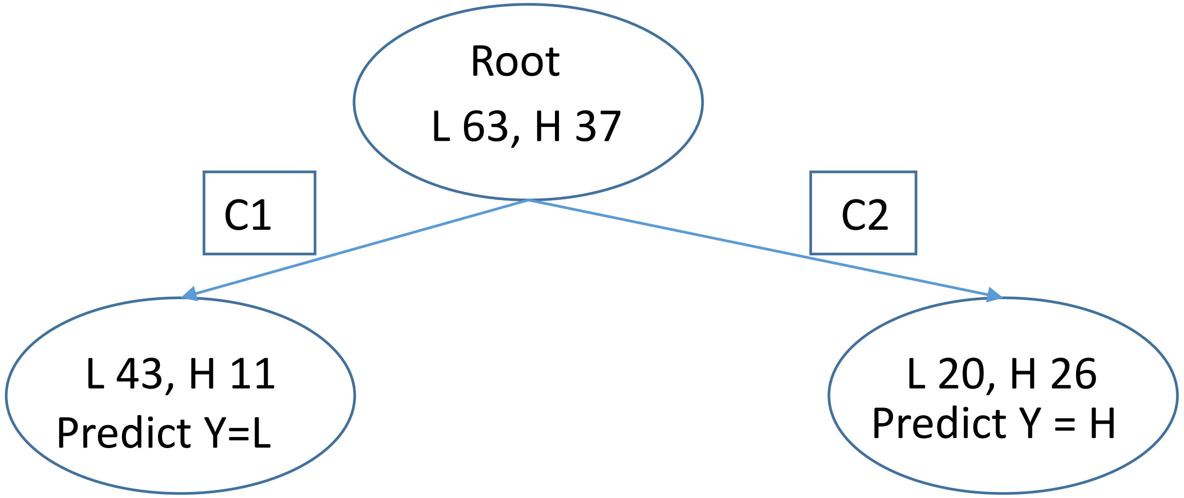
Original Entropy – (Split1_size*Split1_ entropy + Split2_size*Split2_ entropy)

0.66 – (90/100*0.67 + 10/100*0.5) ~ 0.01!!

Feature selection based on entropy

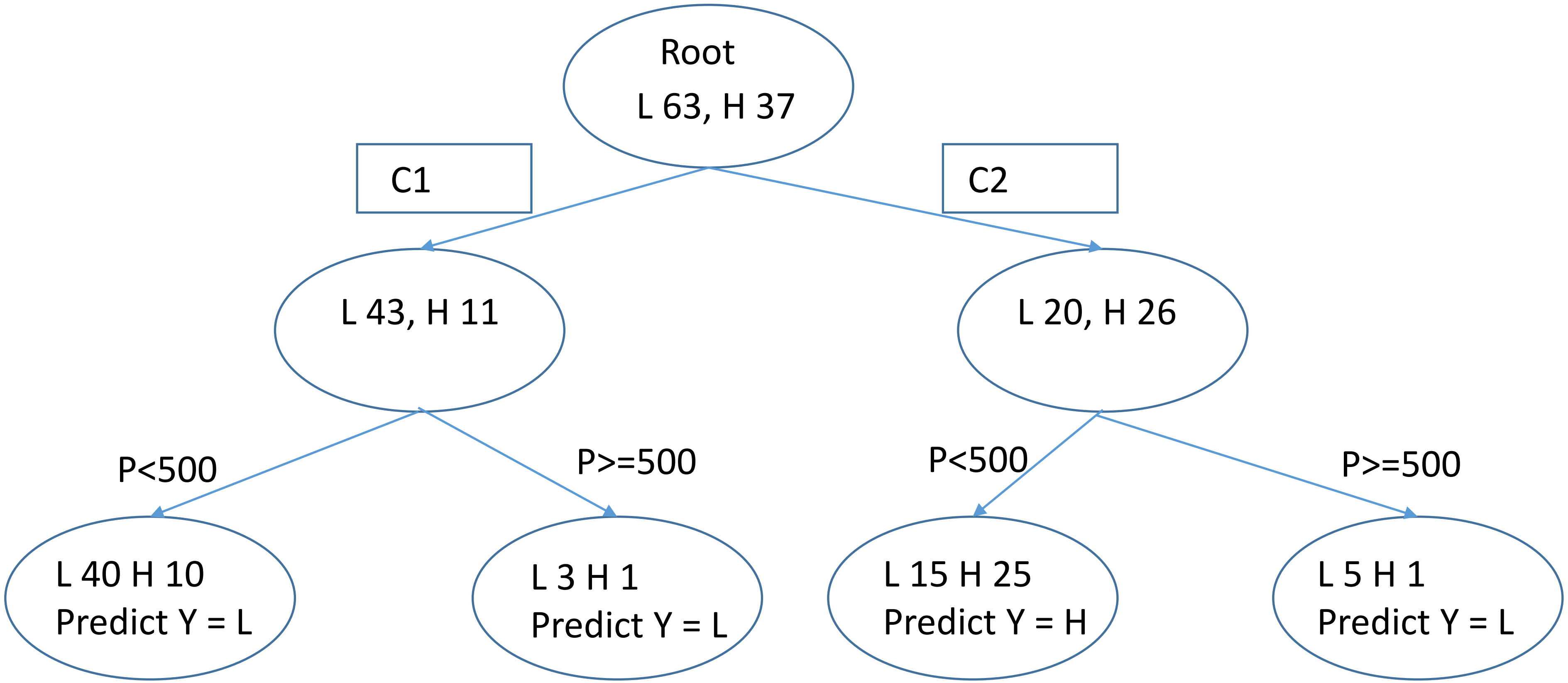
- Each discrete feature splits the dataset
- Continuous features can always be converted to discrete
- “Pure” dataset: - disbalanced class distribution
  - low entropy
  - high information gain
- Choose that feature which provides most information gain!

Decision Stump



Training accuracy: 43/63 for LOW, 26/37 for HIGH, 69/100 OVERALL

Decision Tree



- Does this split provide "information gain"???
- If yes, split. If no, stop at previous step

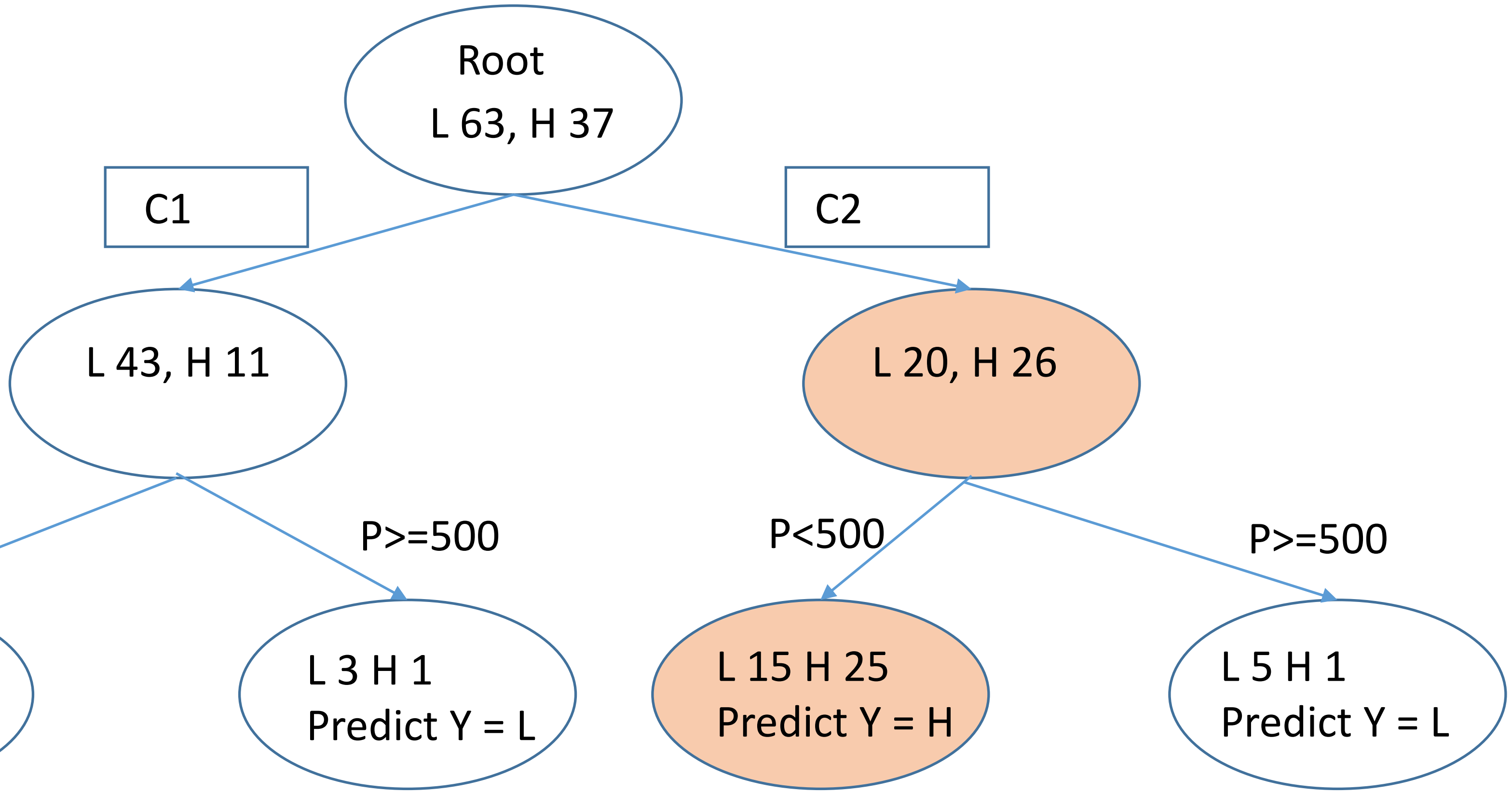
Decision Tree algorithm

- 1. Identify the feature that results in maximum information gain
- 2. Split the dataset accordingly
- 3. Identify if any feature can result in further information gain on the split sets
- 4. If yes, split further. If no, stop.
- 5. Goto 3
- 6. At each leaf, the prediction is the mode label

- Test:
- Follow the sequence of decisions based on the features of test example
- Make prediction according to leaf

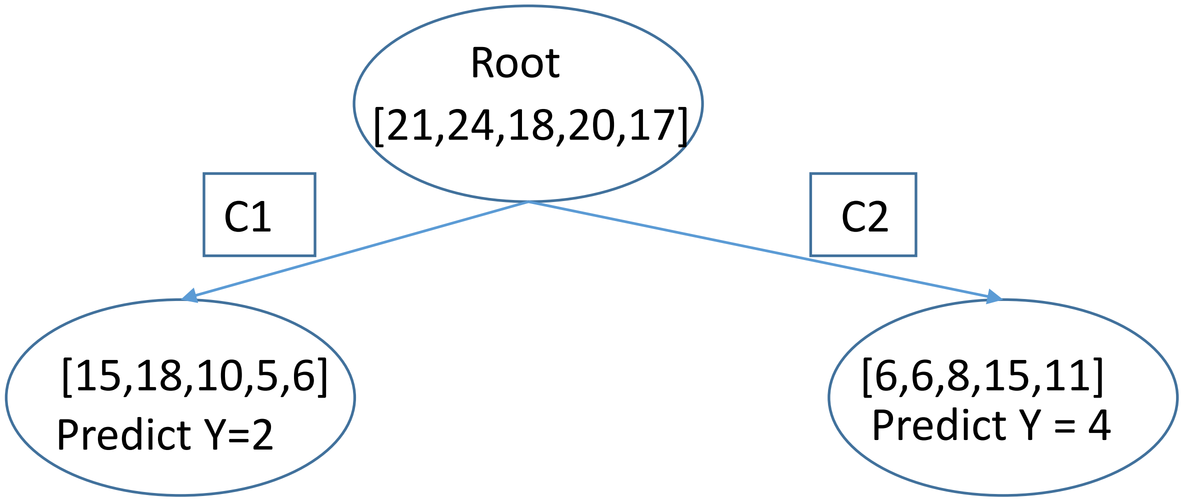
Decision Tree

TEST CASE: COMPANY = C2, PRICE = 350



- Prediction: Y= H

Multi-class Decision Stump



Training accuracy: 18/24 for CLASS 2, 15/20 for CLASS 4, 33/100 OVERALL

Advantages and Disadvantages

Advantage:

- Easy to interpret
- Easy to classify at test time
- Provides a ranking of features (according to usefulness)

Disadvantages:

- No optimal solution known, IG is just heuristic, can create many small branches
- Can cause overfitting if tree grows deep (need to stop growing)

Regression Trees

- What if we want to predict Average Rating of a product?
- Real number between 1 and 5!
- Decision trees can also be used for regression
- Measure of homogeneity at each node: variance of labels (instead of entropy)
- Split criteria: reduction in total variance (instead of information gain)
- Final prediction: Mean label in the leaf node (instead of mode)

	COMPANY=C1	COMPANY=C2	NO SPLIT
COUNT	54	46	100
MEAN of RATINGS	3.0	4.0	3.46
VARIANCE of RATINGS	1.5	0.5	1.1
Reduction in Variance			1.1-(0.54*1.5+0.46*0.5)= 0.06

	VIDEO RATE <30 fps	Video RATE >=30fps	NO SPLIT
COUNT	70	30	100
MEAN of RATINGS	3.1	4.5	3.46
VARIANCE of RATINGS	1.2	0.4	1.1
Reduction in Variance			1.1-(0.7*1.2+0.3*0.4)= 0.14

Regression Tree