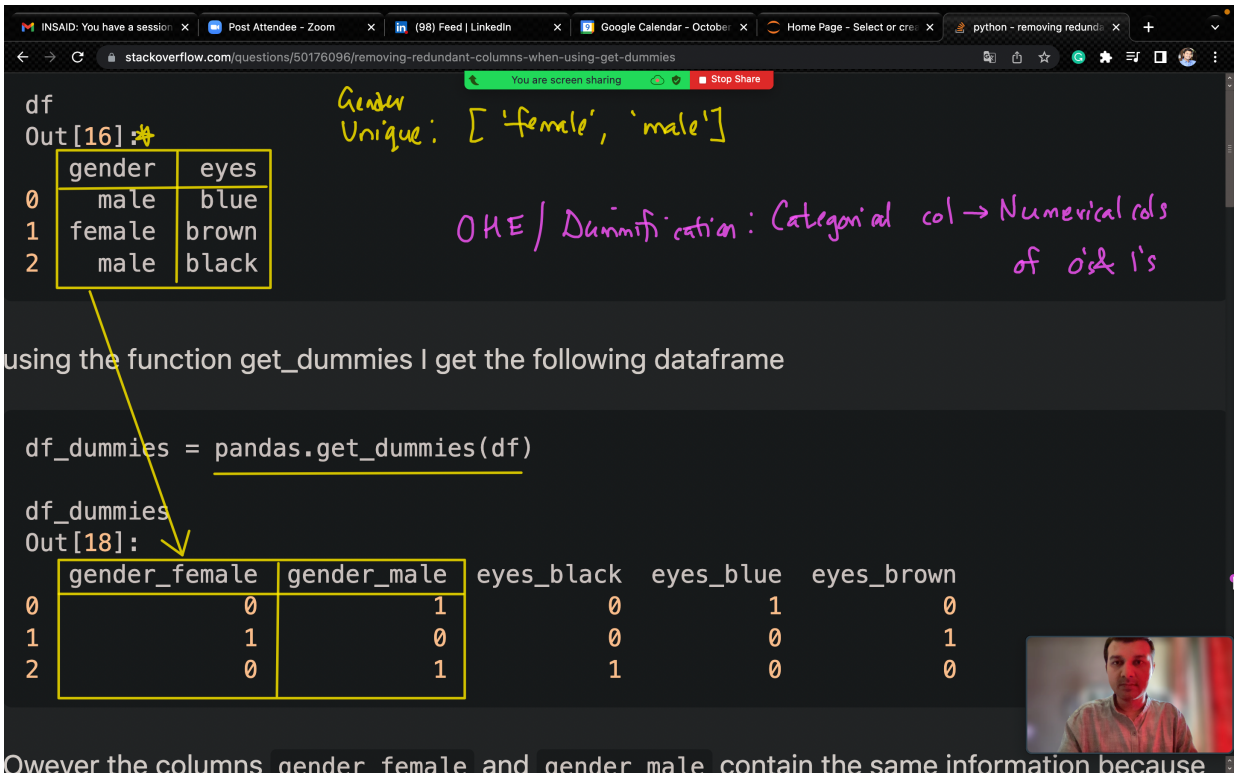
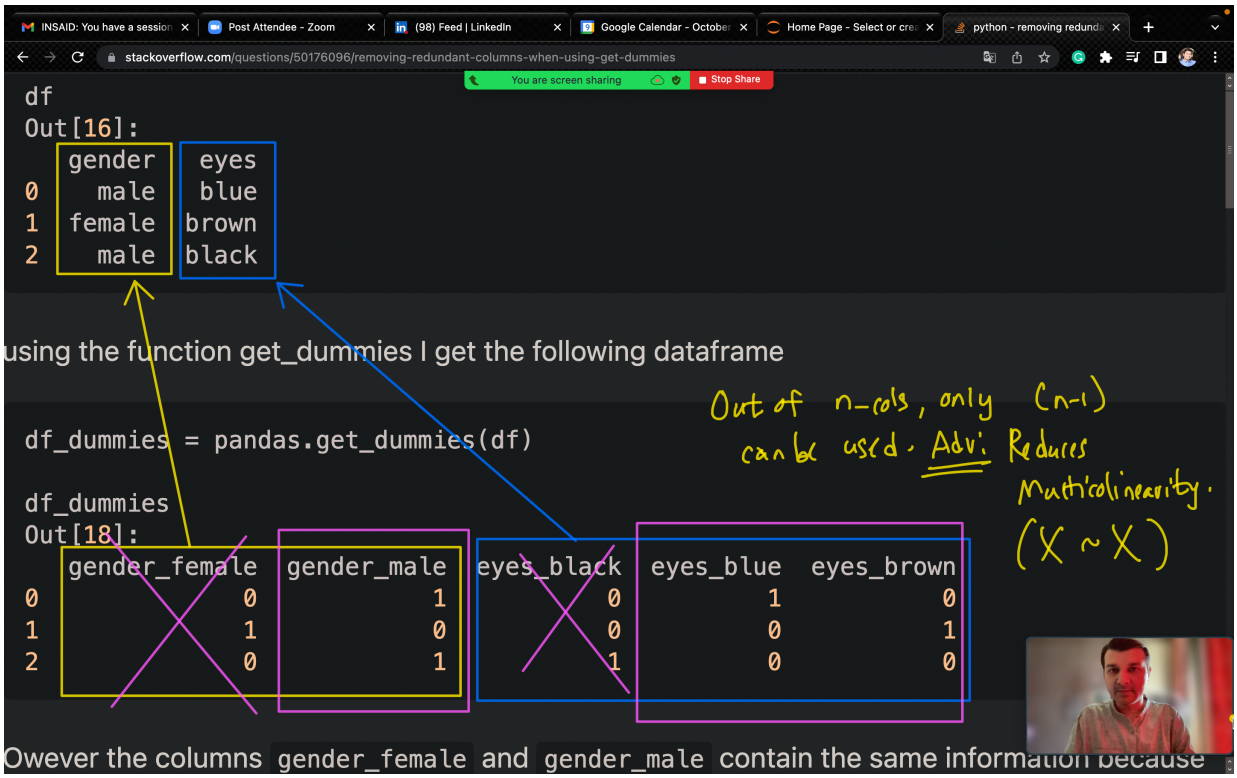
https://stackoverflow.com/questions/50176096/removing-redundant-columns-when-using-get-dummies (https://stackoverflow.com/questions/50176096/removing-redundant-columns-when-using-get-dummies)



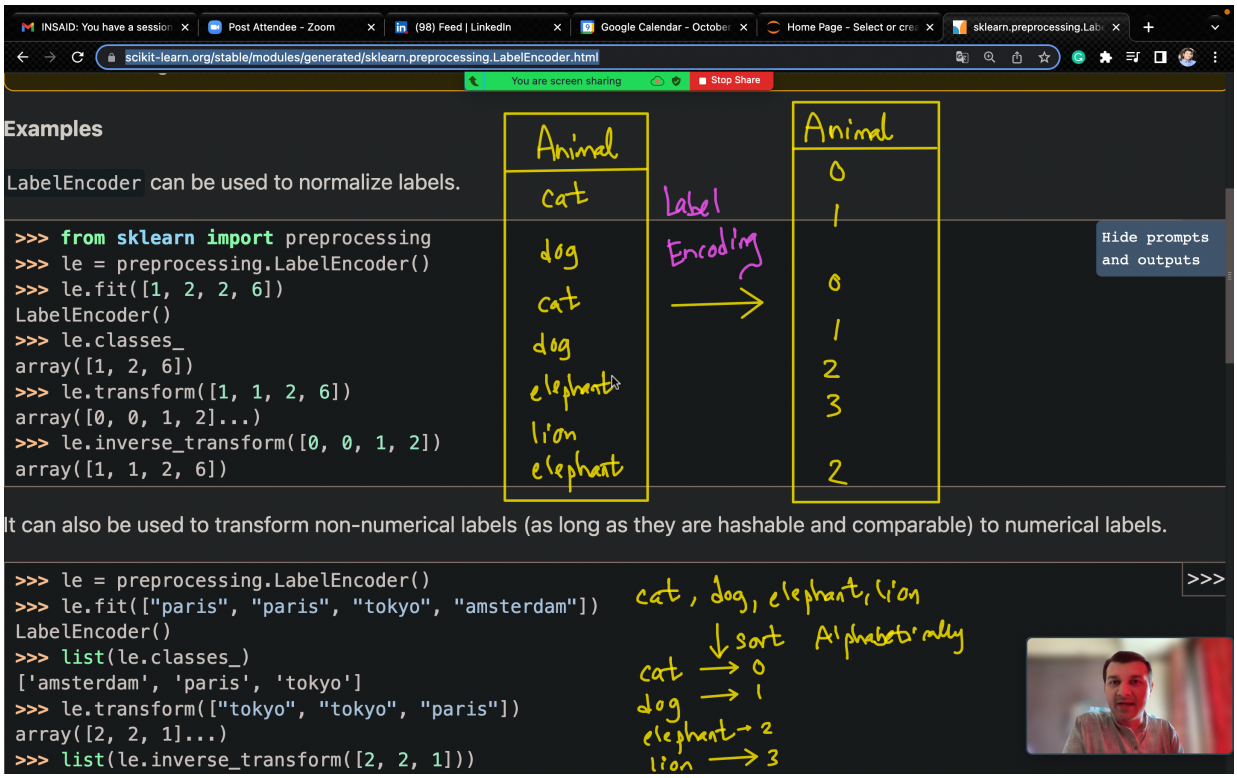
In [ ]:



In [ ]:

https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html (https://scikit-

learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html)



In [1]:
```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

le.fit(["cat","dog","cat","dog","elephant","lion","elephant"])

list(le.classes_)
```

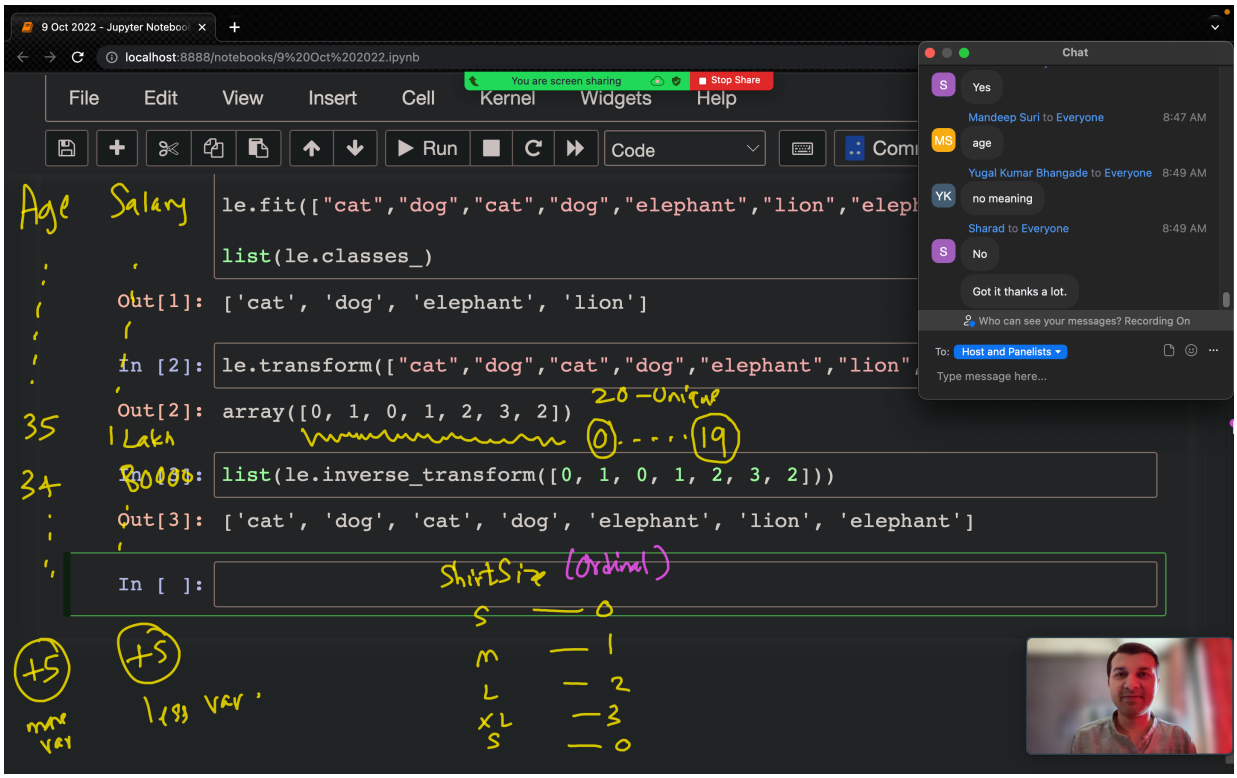
Out[1]:
```
['cat', 'dog', 'elephant', 'lion']
```

In [2]:
```
le.transform(["cat","dog","cat","dog","elephant","lion","elephant"])
```

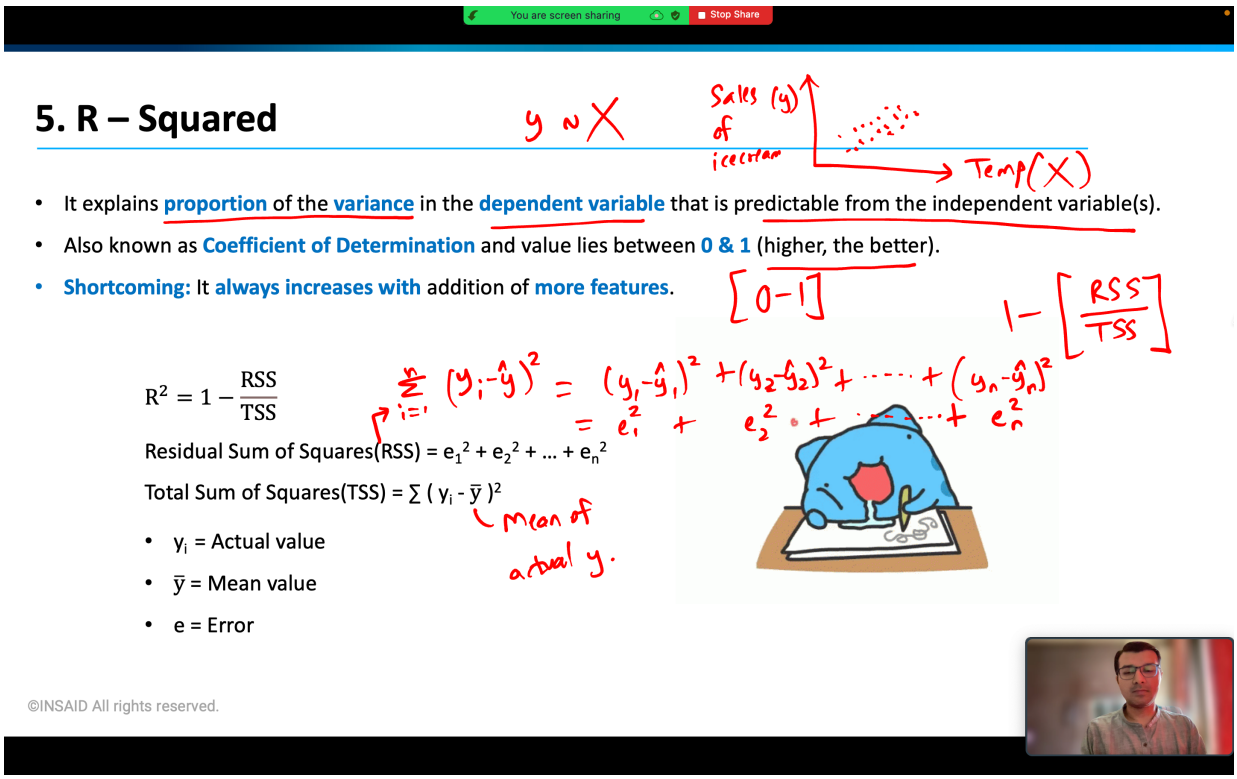
Out[2]:
```
array([0, 1, 0, 1, 2, 3, 2])
```

In [3]:
```
list(le.inverse_transform([0, 1, 0, 1, 2, 3, 2]))
```

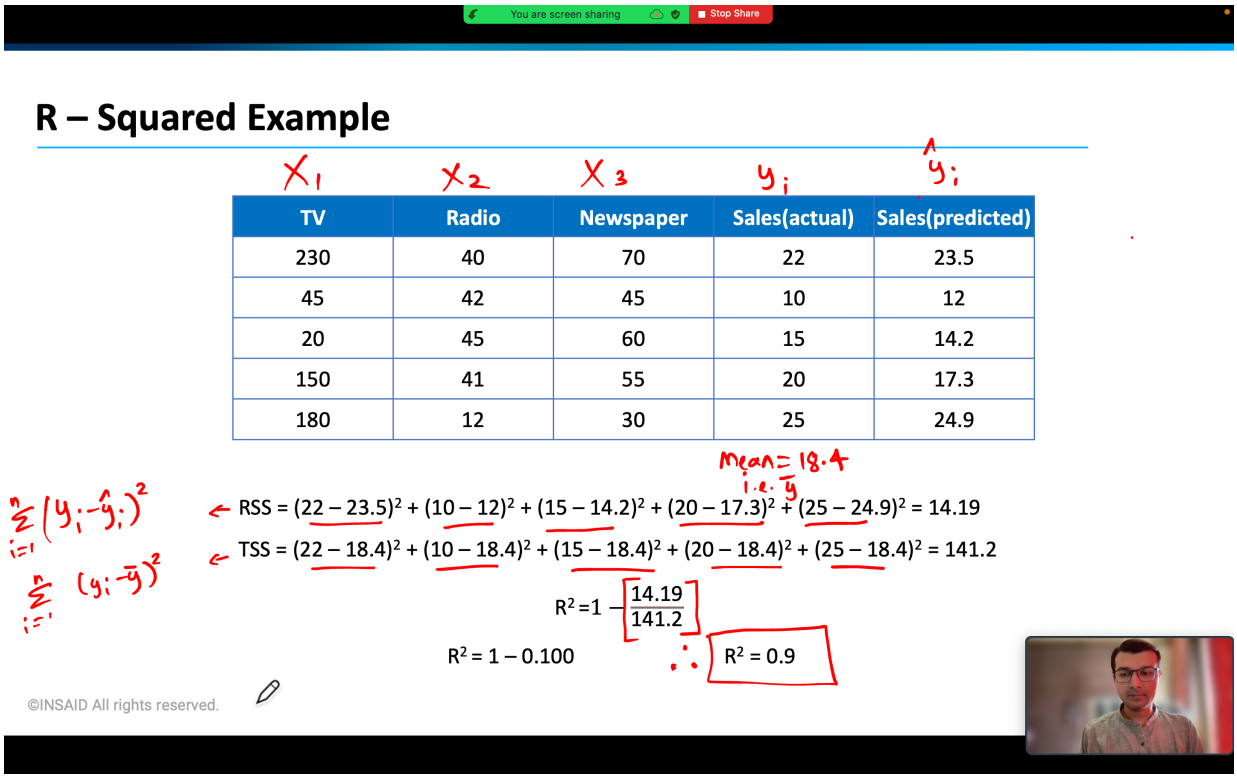
Out[3]:
```
['cat', 'dog', 'cat', 'dog', 'elephant', 'lion', 'elephant']
```


9 Oct 2022 - Jupyter Notebook
localhost:8888/notebooks/9%20Oct%202022.ipynb
You are screen sharing
Stop Share
File Edit View Insert Cell Kernel Widgets Help
Run Code
Chat
Yes
Mandeep Suri to Everyone 8:47 AM
age
Yugal Kumar Bhangade to Everyone 8:49 AM
no meaning
Sharad to Everyone 8:49 AM
No
Got it thanks a lot.
Who can see your messages? Recording On
To: Host and Panelists
Type message here...
Age
Salary
le.fit(["cat","dog","cat","dog","elephant","lion","elep
list(le.classes_)
Out[1]: ['cat', 'dog', 'elephant', 'lion']
In [2]: le.transform(["cat","dog","cat","dog","elephant","lion"
Out[2]: array([0, 1, 0, 1, 2, 3, 2])
20 - Unique
0 ... 19
In [3]: list(le.inverse_transform([0, 1, 0, 1, 2, 3, 2]))
Out[3]: ['cat', 'dog', 'cat', 'dog', 'elephant', 'lion', 'elephant']
In [ ]:
ShirtSize (Ordinal)
S — 0
M — 1
L — 2
XL — 3
S — 0
+5
+5
1 less var
more var

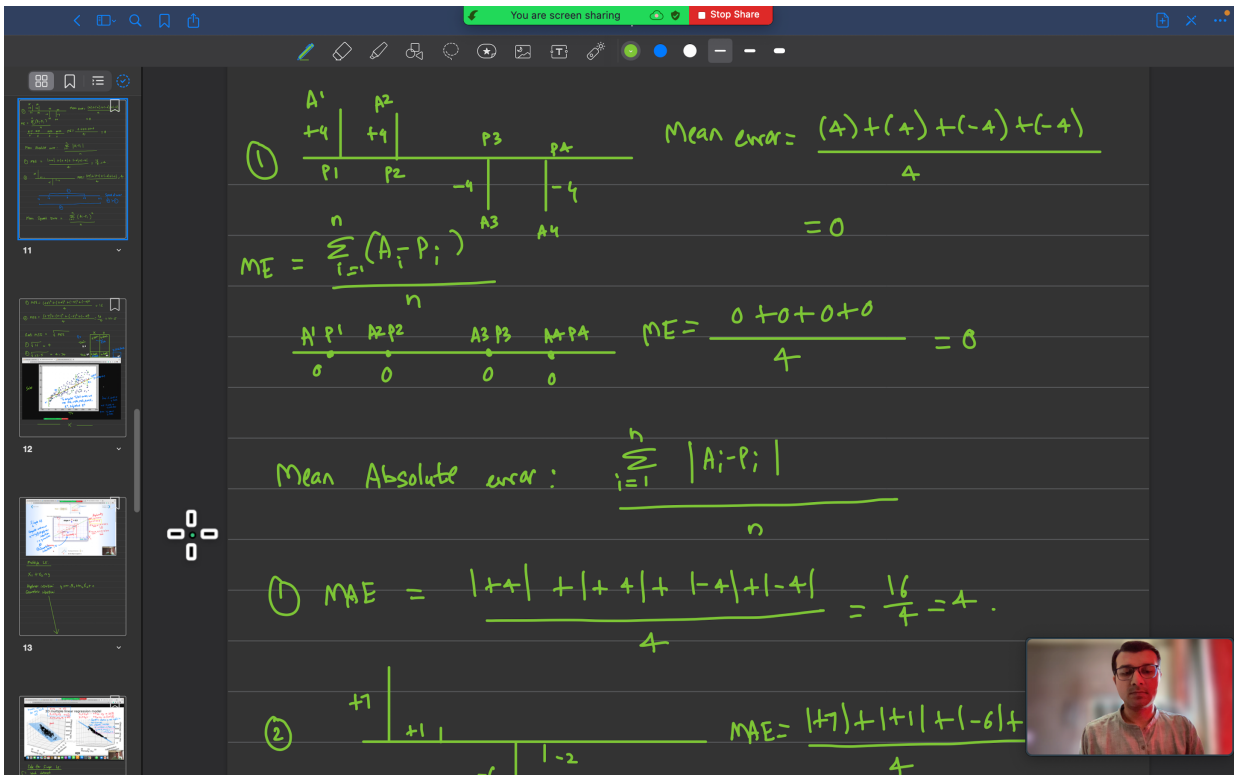
In [ ]:


You are screen sharing
Stop Share
5. R – Squared
y ∝ X
Sales (y) of icecream
Temp(X)
It explains proportion of the variance in the dependent variable that is predictable from the independent variable(s).
Also known as Coefficient of Determination and value lies between 0 & 1 (higher, the better).
Shortcoming: It always increases with addition of more features.
[0-1]
1 - [RSS/TSS]
$R^2 = 1 - \frac{RSS}{TSS}$
$\sum_{i=1}^{n} (y_i - \hat{y})^2 = (y_1-\hat{y}_1)^2 + (y_2-\hat{y}_2)^2 + \cdots + (y_n-\hat{y}_n)^2$
$= e_1^2 + e_2^2 + \cdots + e_n^2$
Residual Sum of Squares(RSS) = $e_1^2 + e_2^2 + \ldots + e_n^2$
Total Sum of Squares(TSS) = $\sum (y_i - \bar{y})^2$
Mean of actual y.
y_i = Actual value
$\bar{y}$ = Mean value
e = Error
©INSAID All rights reserved.

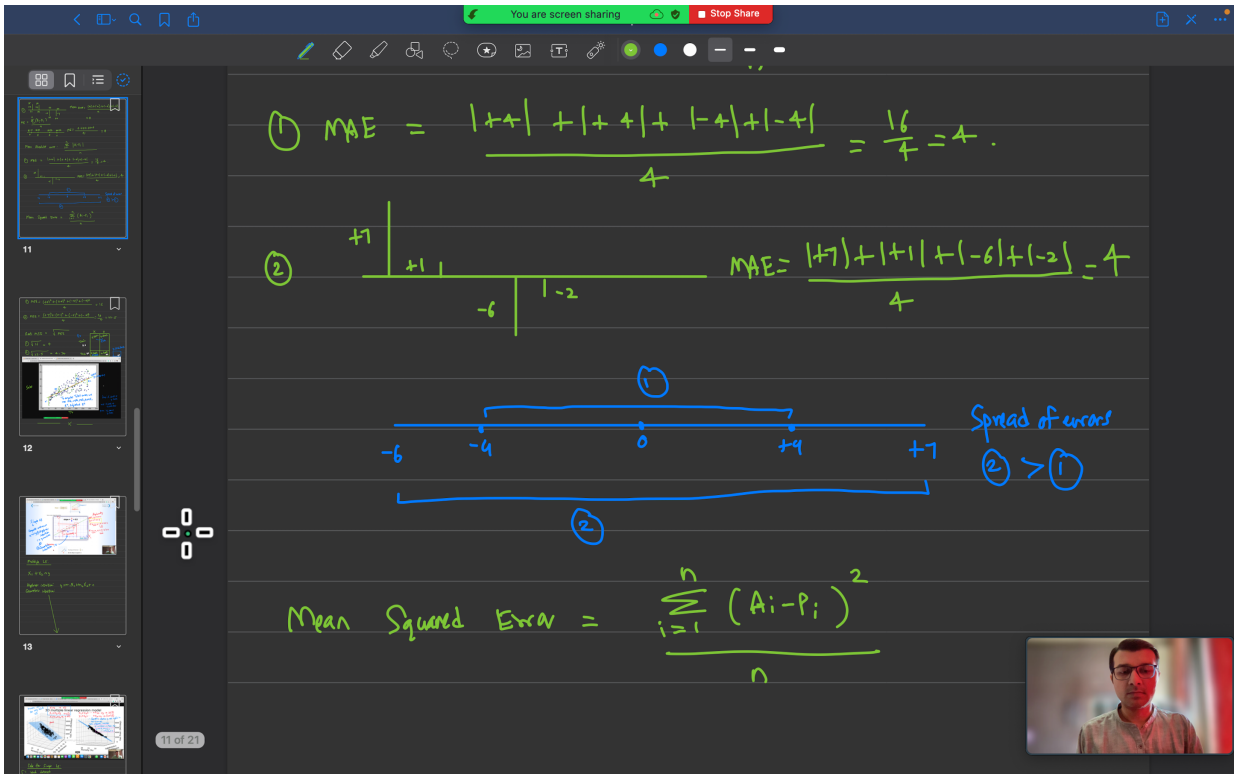
In [ ]:


You are screen sharing
Stop Share
R – Squared Example
X1 X2 X3 yi ŷi
TV | Radio | Newspaper | Sales(actual) | Sales(predicted)
230 | 40 | 70 | 22 | 23.5
45 | 42 | 45 | 10 | 12
20 | 45 | 60 | 15 | 14.2
150 | 41 | 55 | 20 | 17.3
180 | 12 | 30 | 25 | 24.9
Mean = 18.4 i.e. ȳ
RSS = (22 – 23.5)² + (10 – 12)² + (15 – 14.2)² + (20 – 17.3)² + (25 – 24.9)² = 14.19
TSS = (22 – 18.4)² + (10 – 18.4)² + (15 – 18.4)² + (20 – 18.4)² + (25 – 18.4)² = 141.2
R² = 1 – 14.19/141.2
R² = 1 – 0.100
∴ R² = 0.9
©INSAID All rights reserved.

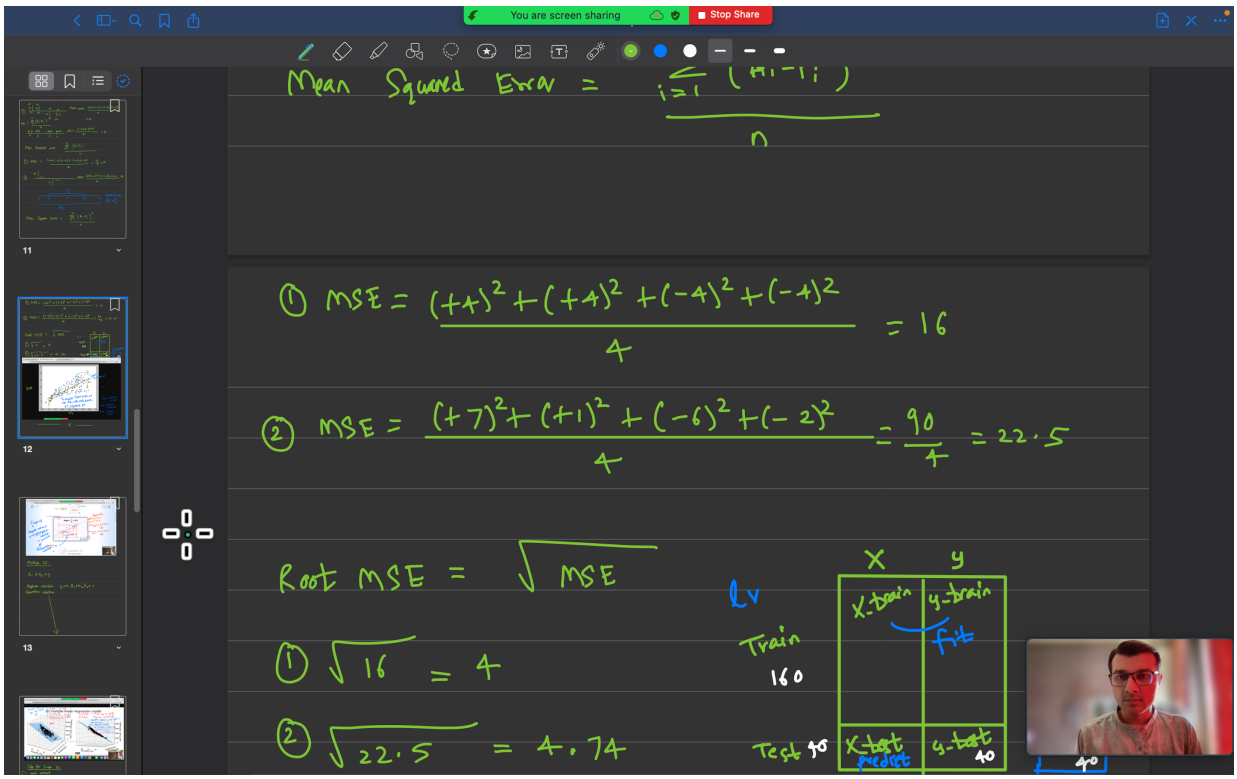
In [ ]:


You are screen sharing
Stop Share
Mean error = (4)+(4)+(-4)+(-4) / 4 = 0
ME = Σ (Ai – Pi) / n
ME = 0+0+0+0 / 4 = 0
Mean Absolute error : Σ |Ai – Pi| / n
MAE = |+4| + |+4| + |-4| + |-4| / 4 = 16/4 = 4.
MAE = |+7| + |+1| + |-6| + ...

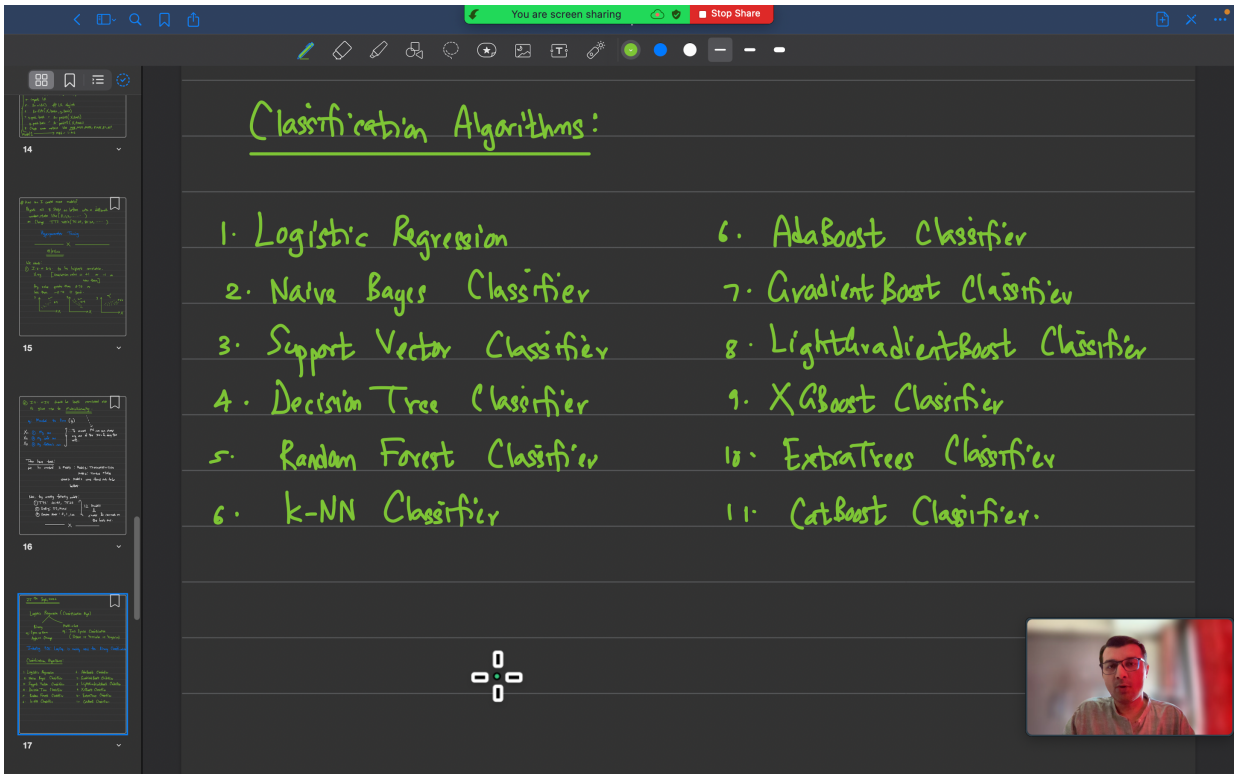
In [ ]:

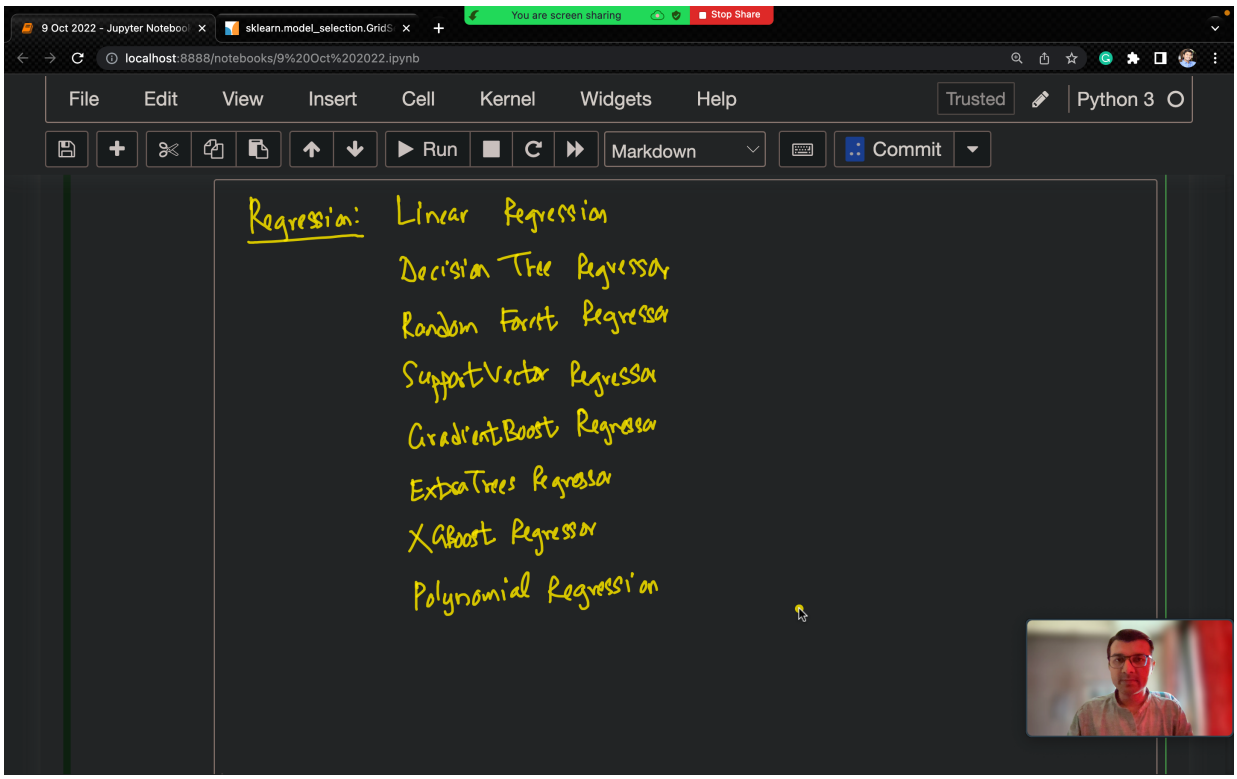

You are screen sharing
Stop Share
① MAE = |+4| + |+4| + |-4| + |-4| / 4 = 16/4 = 4.
②
+7
+1
-6
-2
MAE = |+7| + |+1| + |-6| + |-2| / 4 = 4
-6
-4
0
+4
+7
Spread of errors
② > ①
Mean Squared Error = Σ (Ai - Pi)^2 / n
11 of 21

In [ ]:

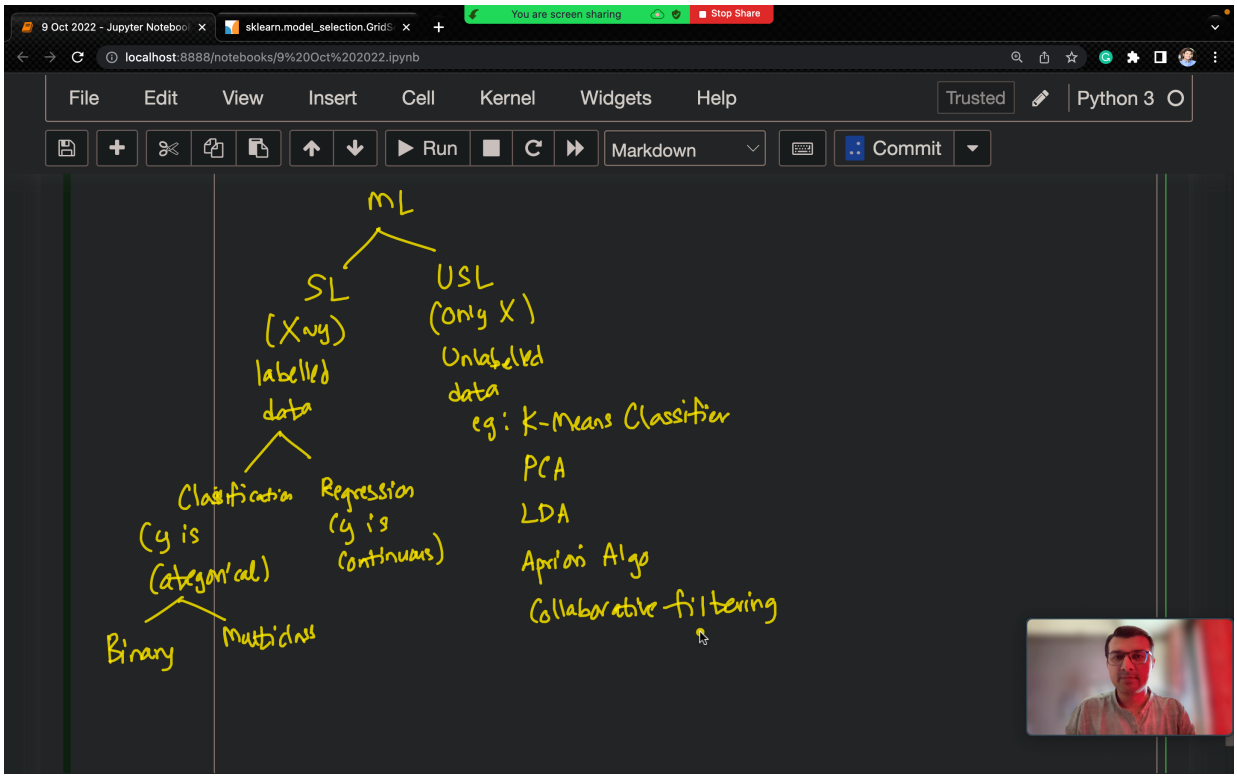

You are screen sharing
Stop Share
Mean Squared Error = Σ (Ai - Pi)^2 / n
① MSE = (+4)^2 + (+4)^2 + (-4)^2 + (-4)^2 / 4 = 16
② MSE = (+7)^2 + (+1)^2 + (-6)^2 + (-2)^2 / 4 = 90/4 = 22.5
Root MSE = √MSE
① √16 = 4
② √22.5 = 4.74
X
y
X_train
y_train
fit
Train
160
Test 40
X_test predict
y_test
40
40

In [ ]:


You are screen sharing
Stop Share
Classification Algorithms:
1. Logistic Regression
2. Naive Bayes Classifier
3. Support Vector Classifier
4. Decision Tree Classifier
5. Random Forest Classifier
6. k-NN Classifier
6. AdaBoost Classifier
7. Gradient Boost Classifier
8. LightGradientBoost Classifier
9. XGBoost Classifier
10. ExtraTrees Classifier
11. CatBoost Classifier.


You are screen sharing
Stop Share
9 Oct 2022 - Jupyter Notebook
sklearn.model_selection.GridS
localhost:8888/notebooks/9%20Oct%202022.ipynb
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3
Run
Markdown
Commit
Regression: Linear Regression
Decision Tree Regressor
Random Forest Regressor
SupportVector Regressor
GradientBoost Regressor
ExtraTrees Regressor
XGBoost Regressor
Polynomial Regression

In [ ]:



In [ ]:

Happy Learning