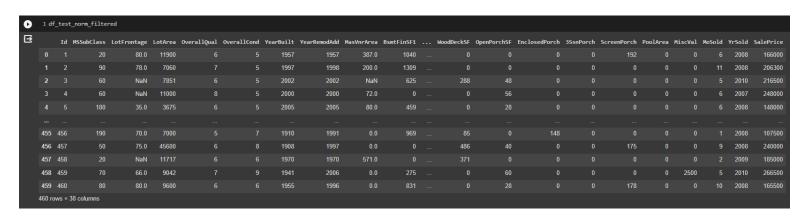
Task 1
Normalized Data Attributes

Train

1 df_tra	in_norm_filte	red																	
Ic	MSSubClass	LotFrontage	LotArea	OverallQual	OverallCond	YearBuilt	YearRemodAdd	MasVnrArea	BsmtFinSF1	WoodDeckSF	OpenPorchSF	EnclosedPorch	3SsnPorch	ScreenPorch	PoolArea	MiscVal	MoSold	YrSold	SalePrice
	120	30.0	9549			1995	1996	0.0	437		30			216				2006	270000
	! 45	60.0	9000			1928	1950	0.0				91					10	2009	76000
		NaN	12394			2003	2003			100								2007	225000
	20	75.0	11957			2006	2006	53.0	24	144	104							2008	232000
		NaN	11616			1962	1962	116.0				144						2009	139000
995 996		50.0	6000			1936	1950											2009	128000
996 997	20	60.0	6960			1970	1970	0.0	375	96						500		2009	120500
997 998		78.0	10206			2007	2007	468.0		144								2008	245000
998 999	60	NaN	11214			1998	1999	0.0		89								2006	199900
999 1000	190	75.0	11625			1965	1965		841									2010	131500
000 rows :	38 columns																		

Test



Selected Features: ['OverallQual', 'GrLivArea', 'BsmtFinSF1', 'GarageArea', 'YearRemodAdd', 'MSSubClass', 'MasVnrArea', 'Fireplaces', 'TotalBsmtSF', 'YearBuilt']

Sample Train data with selected features

1 X_	_train									
	OverallQual	GrLivArea	BsmtFinSF1	GarageArea	YearRemodAdd	MSSubClass	MasVnrArea	Fireplaces	TotalBsmtSF	YearBuilt
0	0.045134	0.033232	0.025377	0.032285	0.036310	0.061385	0.000000	0.04222	0.046697	0.036535
1	0.033850	0.017439	0.000000	0.024163	0.035473	0.023019	0.000000	0.00000	0.024505	0.035308
2	0.045134	0.035011	0.001394	0.055307	0.036492	0.010231	0.008637	0.04222	0.049197	0.036737
3	0.039492	0.033921	0.081298	0.036312	0.036383	0.010231	0.039763	0.00000	0.047666	0.036627
4	0.039492	0.036257	0.000000	0.030271	0.036401	0.010231	0.027866	0.04222	0.050948	0.036627
761	0.016925	0.023133	0.000000	0.026848	0.035473	0.046039	0.000000	0.00000	0.000000	0.035711
762	0.022567	0.029450	0.039023	0.016109	0.035473	0.025577	0.000000	0.00000	0.021004	0.035455
763	0.022567	0.019218	0.021776	0.044300	0.035837	0.010231	0.000000	0.00000	0.027005	0.036078
764	0.045134	0.034766	0.001916	0.050877	0.036510	0.010231	0.076266	0.04222	0.048853	0.036755
765	0.028209	0.023111	0.048837	0.033829	0.035746	0.097193	0.000000	0.00000	0.032475	0.035986
766 rc	ows × 10 column	IS								

```
Mean Squared Error (MSE) Train Error: 3.826873504952882e-05
Mean Squared Error (MSE) Test Error: 0.009589233903404918
```

Task 2

5th degree Polynomial Transform

```
X_train shape before -> after transform (766, 10) --> (766, 3003)
X_test shape before -> after transform (355, 10) --> (355, 3003)
```

Task 3

```
Train Error(MSE) on Polynomially transformed data(deg=5) : 1.0209895058632275e-08
Test Error (MSE) on Polynomially Transformed data(deg=5) : 184.51528357862554
```

As we can see train error is low, but test error is increasing so we can say overfitting has occurred and we need to regularize it to fit model well.

values for the regularization parameter

```
reg_values = [0.001, 0.01, 0.1, 1, 5, 10, 20, 50, 100, 1000]
```

Task 5

Chosen MLP parameters.

```
4 parameters = {
5    'hidden_layer_sizes': [(50,), (100,), (50, 50), (100, 100)],
6    'alpha': [0.0001, 0.001, 0.01],
7 }
```

```
Mean cross-validation Error for all models:

ECV -0.000371012, Parameters: {'alpha': 0.0001, 'hidden_layer_sizes': (50,)}

ECV -0.000195479, Parameters: {'alpha': 0.0001, 'hidden_layer_sizes': (100,)}

ECV -0.000202535, Parameters: {'alpha': 0.0001, 'hidden_layer_sizes': (50, 50)}

ECV -0.000141606, Parameters: {'alpha': 0.0001, 'hidden_layer_sizes': (100, 100)}

ECV -0.000271681, Parameters: {'alpha': 0.001, 'hidden_layer_sizes': (50,)}

ECV -0.000196039, Parameters: {'alpha': 0.001, 'hidden_layer_sizes': (100,)}

ECV -0.000158889, Parameters: {'alpha': 0.001, 'hidden_layer_sizes': (50, 50)}

ECV -0.000113189, Parameters: {'alpha': 0.001, 'hidden_layer_sizes': (100, 100)}

ECV -0.000311873, Parameters: {'alpha': 0.01, 'hidden_layer_sizes': (50,)}

ECV -0.000199820, Parameters: {'alpha': 0.01, 'hidden_layer_sizes': (50,)}

ECV -0.000215858, Parameters: {'alpha': 0.01, 'hidden_layer_sizes': (50, 50)}

ECV -0.000105994, Parameters: {'alpha': 0.01, 'hidden_layer_sizes': (50, 50)}
```

```
Optimal Model Parameters : {'alpha': 0.01, 'hidden_layer_sizes': (100, 100)}

Best cross-validation Error (negative MSE): -0.00010599390884324517

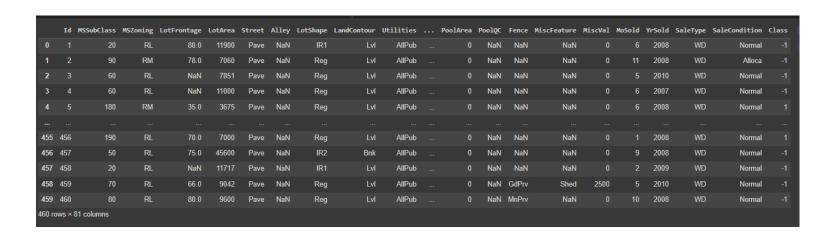
Train MSE with regp 0.01 is : 7.732782172785152e-05

Test MSE with regp 0.01 is: 0.00026736394202103806
```

Task 6
Train data with added Attribute Class (1, -1) based on Sale Price

		_class MSSubClass	MS7oning	LotErontage	LotArea	Street	Allev	LotShane	LandContour	Utilitie	Pool00	Fence	MiscFeature	MiscVal	MoSold	VrSold	SaleTyne	SaleCondition	SalePrice	Class
0	1	120	RL	30.0	9549	Pave	NaN	IR1	Lvl		NaN	NaN	NaN	0			WD	Normal	270000	
1	2	45		60.0	9000	Pave	NaN	Reg	Lvi				NaN				WD	Normal	76000	
2		60		NaN	12394	Pave		IR1	Lvi				NaN				WD	Family	225000	
3	4	20		75.0	11957	Pave	NaN	IR1	Lvl		NaN	NaN	NaN			2008	WD	Normal	232000	
4		20	RL	NaN	11616	Pave	NaN	Reg	Lvl	AllPul	NaN	NaN	NaN			2009	WD	Normal	139000	
995	996	50	RM	50.0	6000	Pave	NaN	Reg	Lvl	AllPul	NaN	NaN	NaN			2009	WD	Normal	128000	
996	997	20	RL	60.0	6960	Pave	NaN	Reg	Lvl	AllPul	NaN	NaN	Shed	500	11	2009	WD	Normal	120500	
997	998	20	RL	78.0	10206	Pave	NaN	Reg	Lvi	AllPul	NaN	NaN	NaN			2008	WD	Normal	245000	
998	999	60	RL	NaN	11214	Pave	NaN	IR1	Lvl	AllPul	NaN	NaN	NaN			2006	WD	Normal	199900	
999	1000	190	RL	75.0	11625	Pave	NaN	Reg	Lvl	AllPul	NaN	NaN	NaN			2010	WD	Normal	131500	
000 re)ws ×	32 columns																		

Test data with added Attribute Class (1, -1) based on Sale Price



Link to Colab:

https://colab.research.google.com/drive/1lhgZP0uGcRMmNj6GKE7wLkO3tgDom4Tt?authuser=0#scrollTo=qojY7chzPXtH&uniqifier=2