Wine Classification of Quality 1 to 10

Professor Linda N. Sellie

Qiyuan Yin, Wayne Wang

Introduction:

The goal of this project is to evaluate the performances of different machine models when used to predict the quality of the wine. To train the model, we obtain a dataset from Kaggle. The dataset contains 1143 samples; for each sample, 11 attributes---fixed acidity, volatile acidity, citric acid, residual sugar percentage, chlorides percentage, free sulfur dioxide percentage, total sulfur dioxide, density, pH value, and sulfates percentage, and alcohol percentage---and the corresponding wine quality is given as an integer in the range [0, 10].

The project utilizes and evaluates the performances of four different models: linear regression, logistic regression, SVM, and neural network.

Preprocessing:

Before any preprocess and training, it was noticed that the data is distributed extremely unbalanced. The majority of the samples are of quality 5 or quality 6 (Figure 1).

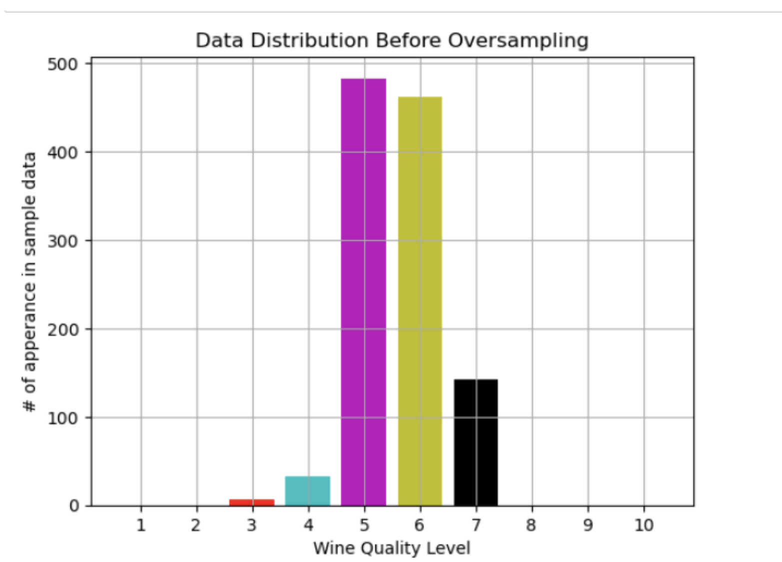

Figure 1: Original Data distribution

To address the imbalance, we use oversampling and exaggerate the number of appearances of other samples. However, oversampling does not fully address the imbalance as there is no wine of quality 1, 2, 8, 9, or 10 in the dataset (Figure 2). This issue does not impact the training process itself, but it will affect the performances of the classification model in a more general setting: models trained based on this dataset do not have the ability to detect wine quality 1, 2, 8, 9, 10. This issue is further addressed in the training process.

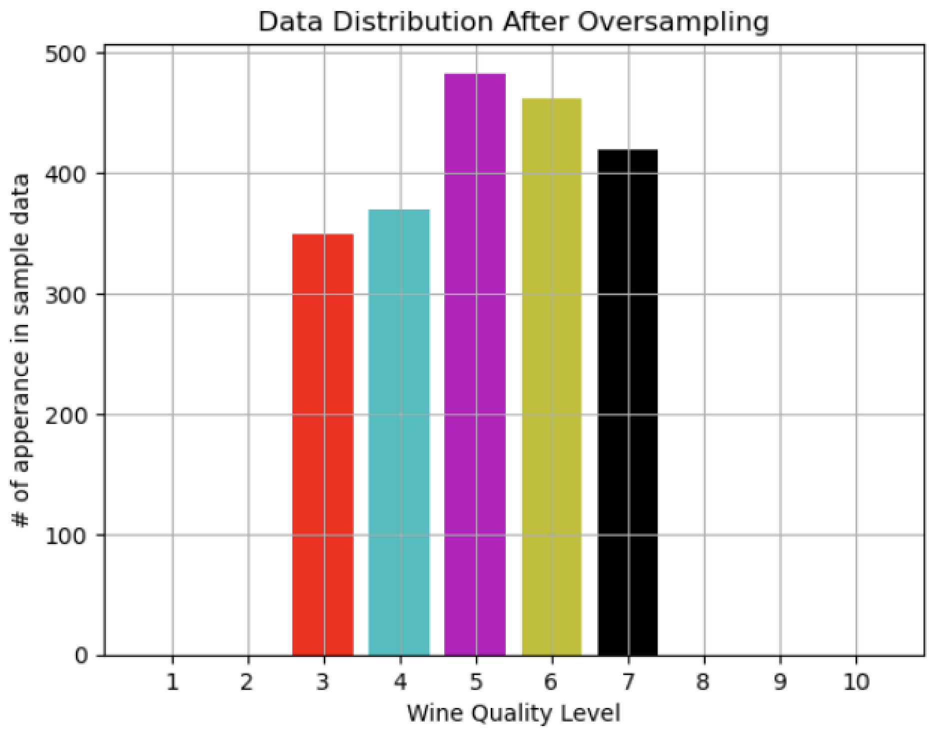

Figure 2: Balanced Data Distribution

Since all of our inputs are numerical, and no data is missing, we scale the data of every attribute using maximum absolute scaling. All our data inputs are greater than 0, after scaling, all our data falls into the range [0,1].

We have divided the dataset into three sets, training set, validation set, and test set. The training set contains 93% of the entire samples, the validation set contains 7% of the training set, and the test set contains 7% of the entire samples.

Supervised Analysis:

SVM

The first model that we examined is SVM. We used sklearn library's SVM module.

Firstly, we examine the performances of different kernel transformations under different regularization parameters of Ridge Regression (L2). We use the linear kernel, RBF kernel, and polynomial kernel.

Before evaluating the performances, we first evaluate the optimized degree of polynomial transformation. To do this, we use the default regularization parameter (C=1.0) and alter the degree of polynomial kernel from 2 to 12 (Figure 3). We have decided to use degree 7 because it has an acceptable validation precision, and it is not overfitting as its validation precision is close to its training precision.

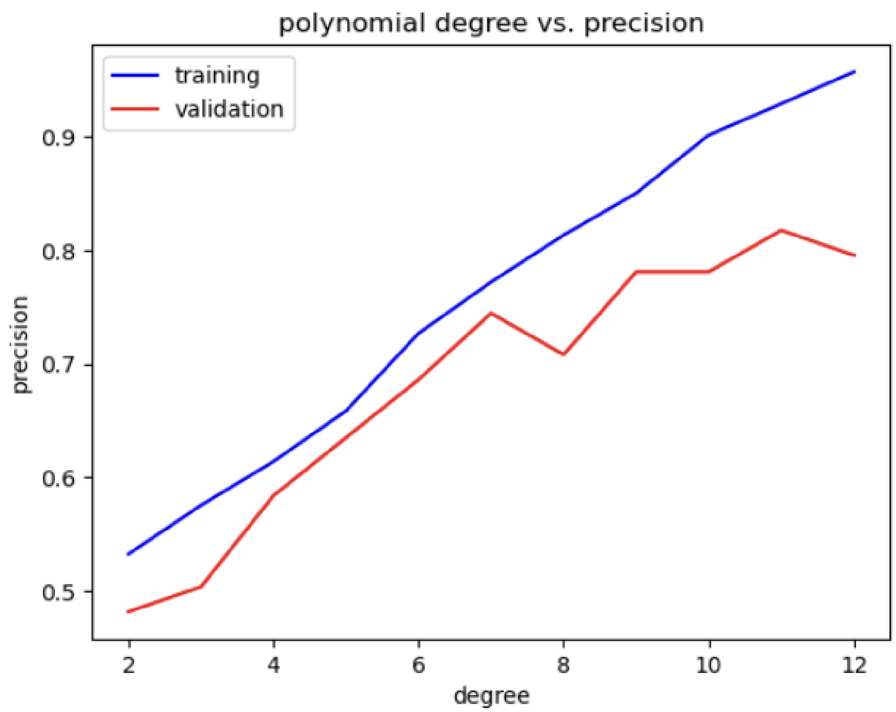

Figure 3: Different Polynomial's Precision (C=1.0)

Then, we try regularizations from set {0.001, 0.005, 0.1, 0.5, 1, 5, 10, 15, 20, 30, 50, 80} on RBF kernel, polynomial kernel of degree 7, and linear kernel.

The result precision of using the polynomial kernel of degree 7 is shown below (Figure 4).

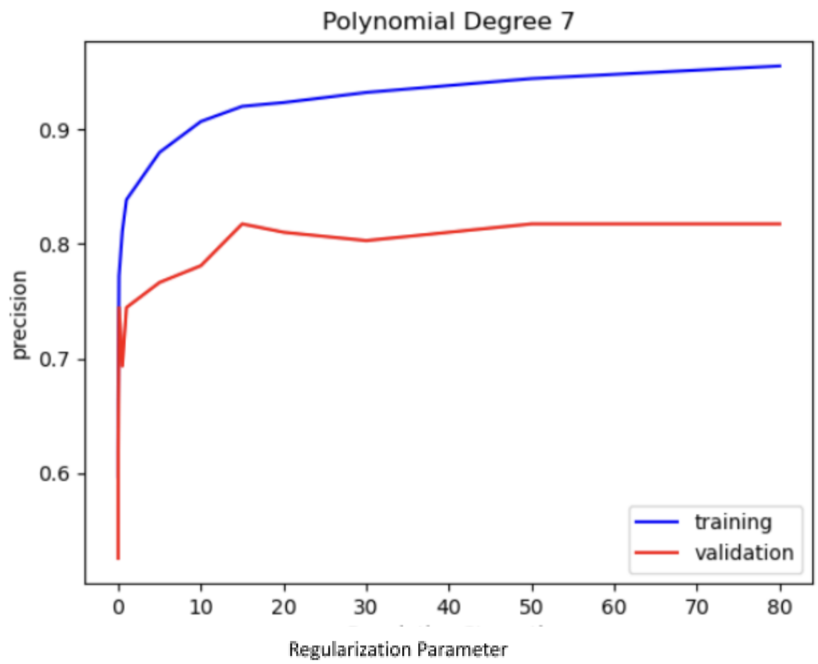

Figure 4: Polynomial Degree 7 – Precision vs. Regularization Parameter

The result precision of using the RBF kernel is shown below (Figure 5).

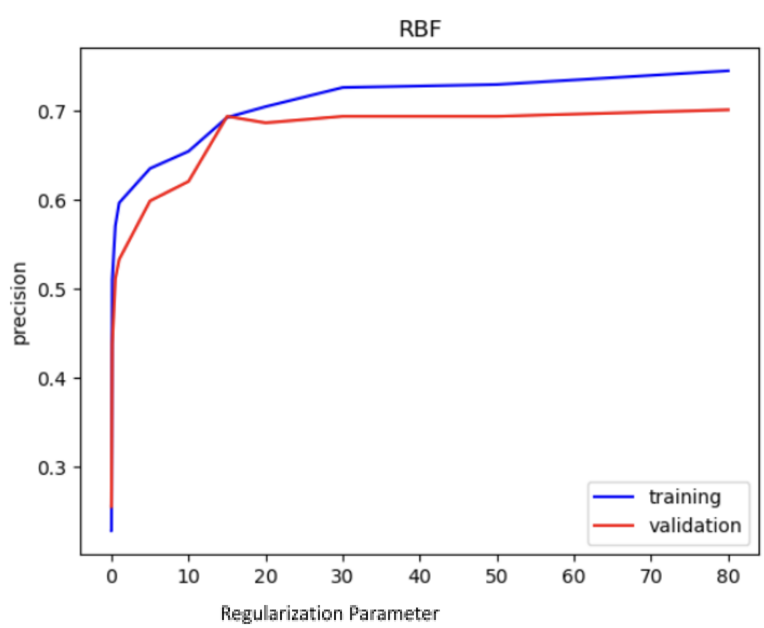

Figure 5: RBF – Precision vs. Regularization Parameter

The result precision of using the linear kernel is shown below (Figure 6).

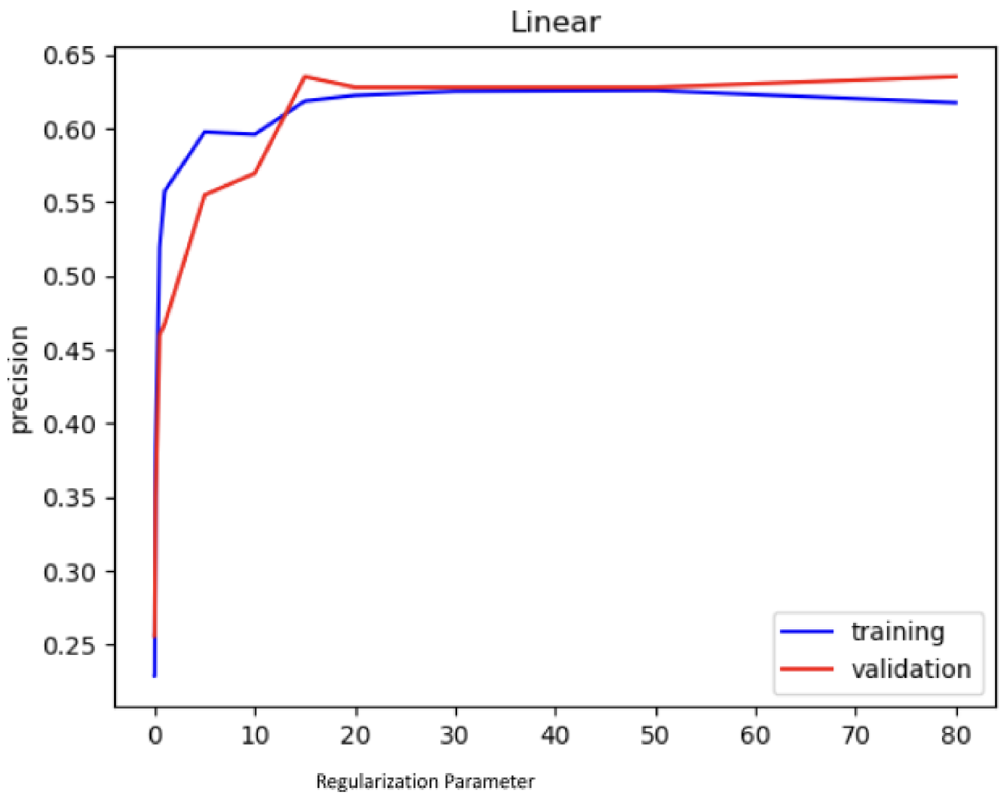

Figure 6: Linear – Precision vs. Regularization Parameter

It is shown that a model using the polynomial degree of 7 and a regularization parameter of 15 has the best validation performances.

Since sklearn does not support showing feature weights when using a non-linear kernel, we will only demonstrate the feature weight changes when using a linear kernel (Figure 7).

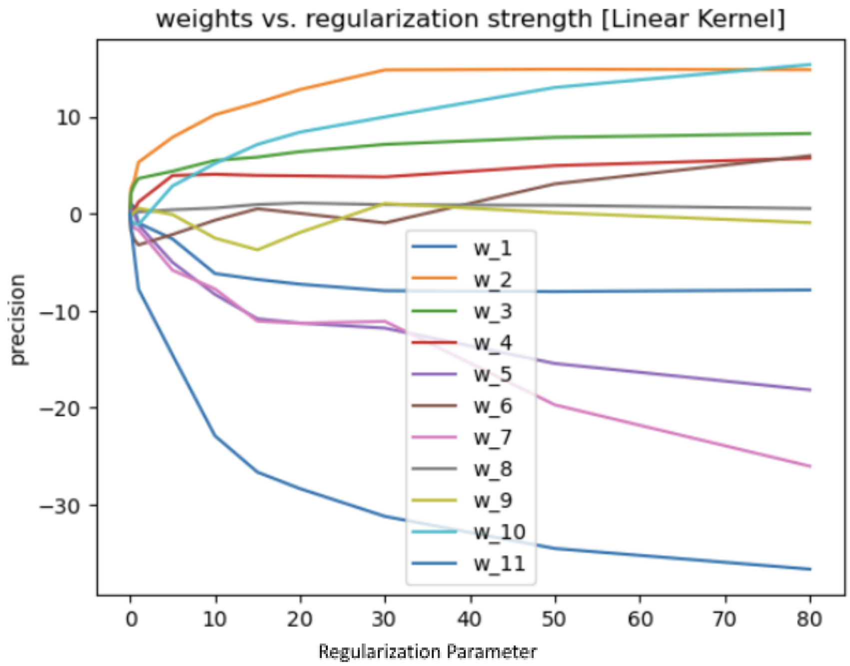

Figure 7: Linear Kernel Feature Weight Changes

Then, we use this model to predict the test set and record the result scores. The results are later used to compare with other optimized models.

Neural Network

The second model that we chose is neural network. For neural networks, we first evaluate which activation should be used. Then, we test different regularization parameters for L2 regularization with different structures of neural networks.

Firstly, we evaluate the performances of sigmoid, tanh, and ReLu activations without tuning any other hyperparameter. The result is shown below (Figure 8). According to the result, ReLu is the best-performed activation function for it has the highest training and validation scores while keeping those two scores very close.

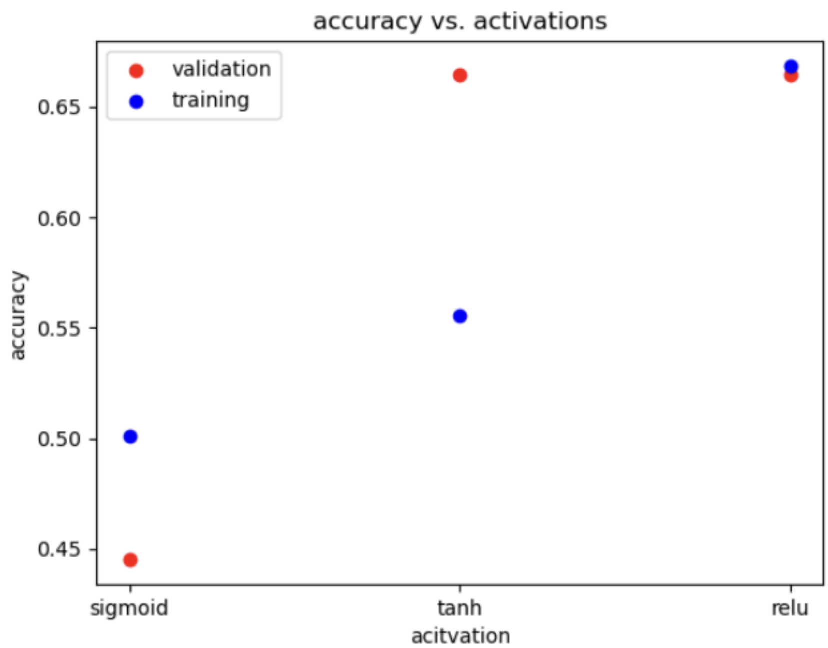

Figure 8: Activation Functions' Performances

Then, we test four neural networks with the following structures for hidden layers: (300), (200, 40), (500, 100, 50), and (500, 200, 100, 30). For each structure, we test the regularization parameters in set C = {0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 1, 5, 10, 20, 40, 60, 75, 100}.

The result of a 4-hidden-layer neural network is shown below (Figure 9).

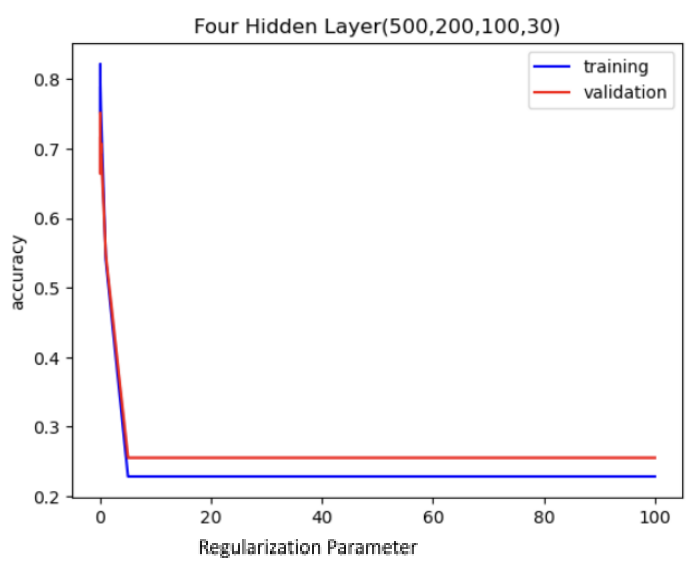

Figure 9: Four Hidden Layer Result

The result of a 3-hidden-layer neural network is shown below (Figure 10).

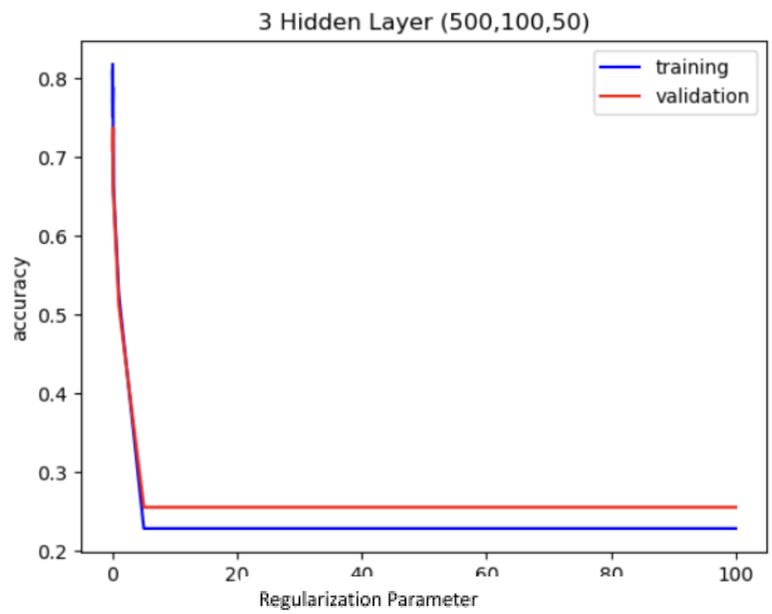

Figure 10: Three Hidden Layer Result

The result of a 2-hidden-layer neural network is shown below (Figure 11).

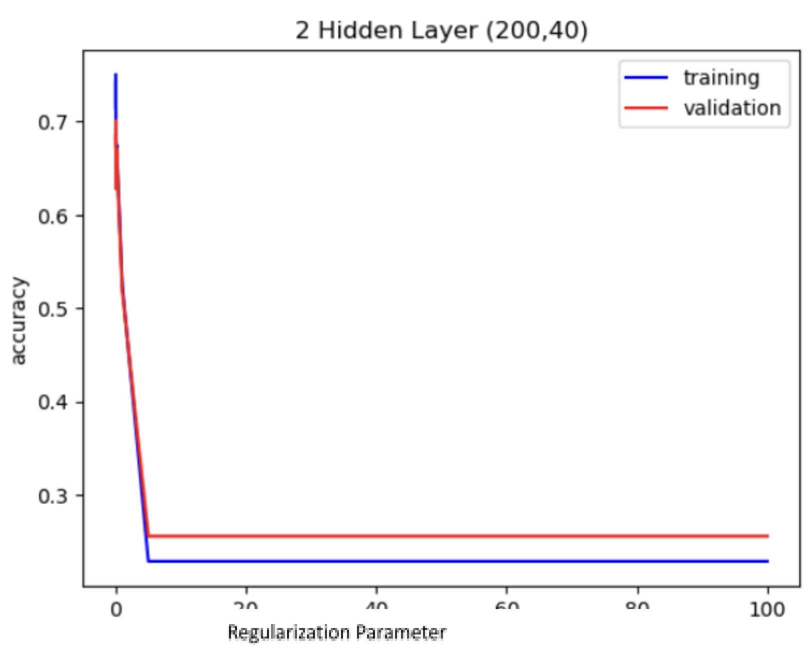

Figure 11: Two Hidden Layer Result

The result of a 1-hidden-layer neural network is shown below (Figure 12).

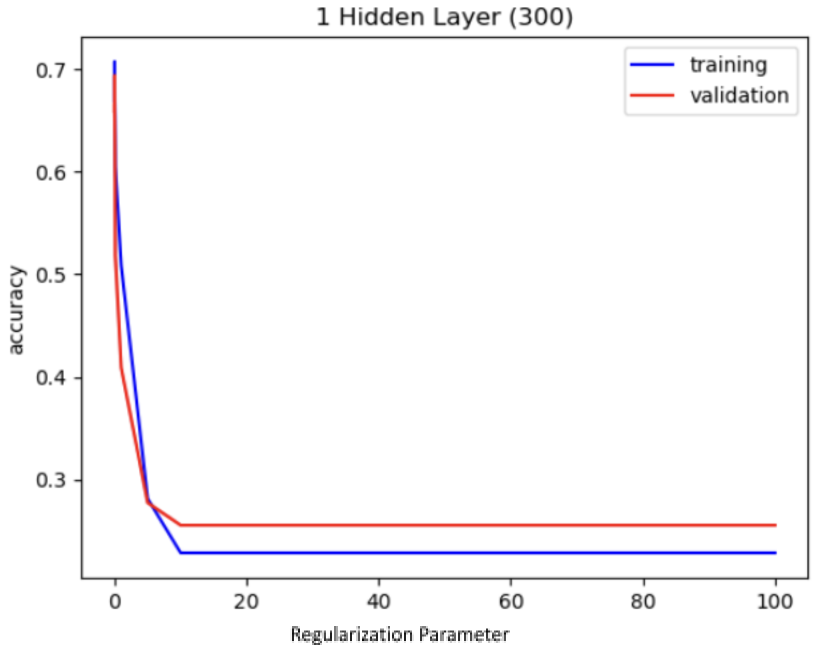

Figure 12: 1 Hidden Layer Result

We observe the more complicated the neural network is, the higher the training and validation scores is; also, the pattern is that the smaller the regularization parameter is, the greater the training and validation scores. To further investigate the relationship between the regularization parameter and the number of hidden layers, we decided to add two extra structures, 5-layer (500,250,125,70,35) and 7-layer (700,500,300,200,150,100,50) to see the performances of the neural network.

The result of a 5-hidden-layer neural network is shown below (Figure 13).

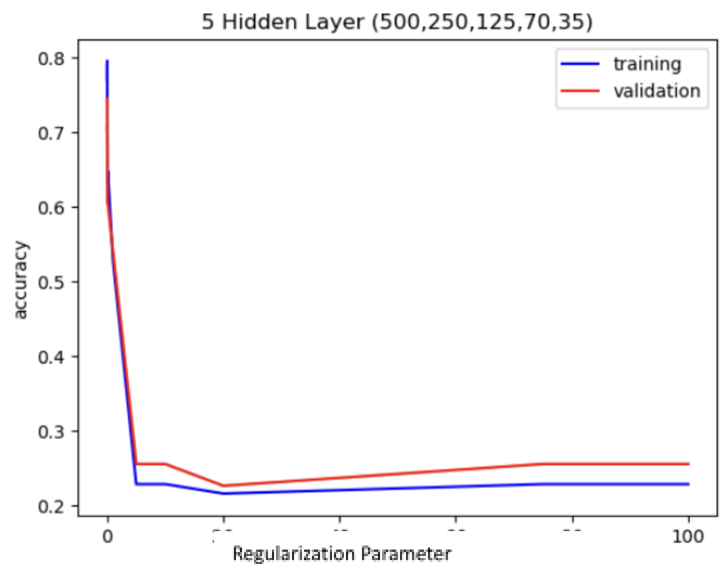

Figure 13: 5 Hidden Layer Result

The result of a 7-hidden-layer neural network is shown below (Figure 14).

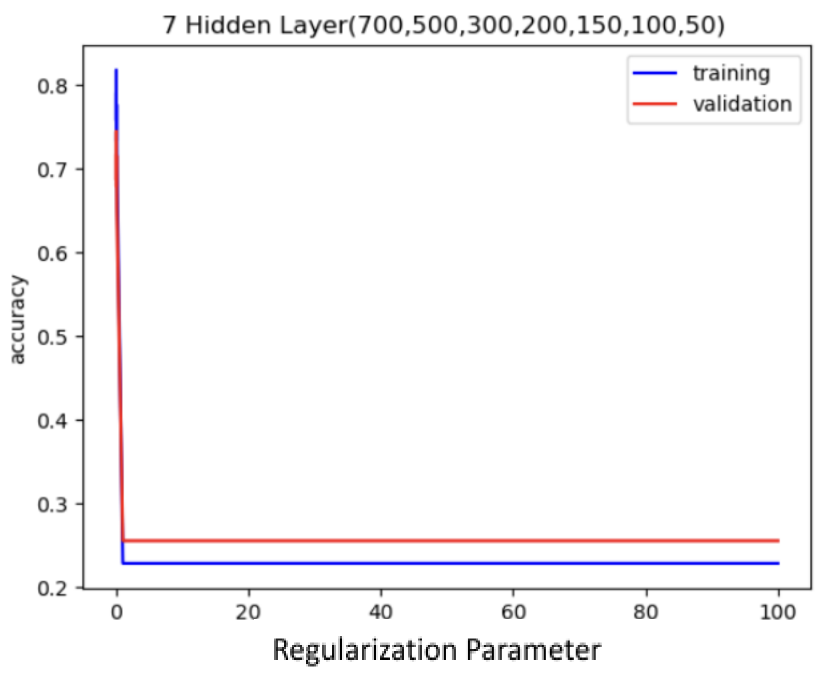

Figure 14: 7 Hidden Layer Result

We have observed that in terms of validation accuracy and the difference between validation accuracy and training accuracy, a structure with 4 hidden layer and a regularization parameter of 0.005 is the optimized model.

Then, we use this model to predict the test set and record the result scores. The results are later used to compare with other optimized models.

Logistic Regression

The third and last classification model that we use is logistic regression.

Firstly, we evaluate the performances of different polynomial transformations in untuned settings. We examine the performances of transformation between degree 1 to degree 7. The result is shown below (Figure 15). We conclude that a polynomial transformation of degree 7 is the best feature transformation as it grants

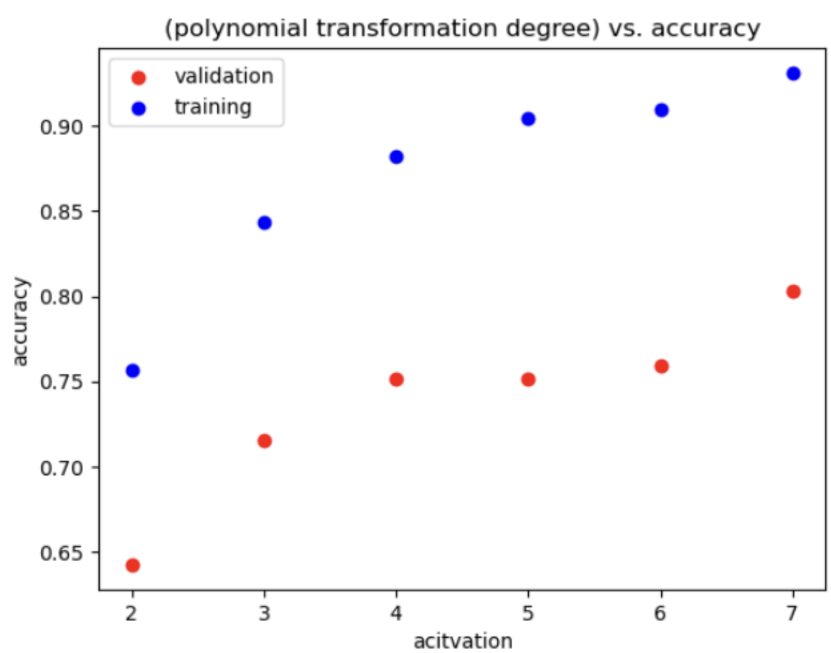

Figure 15: Polynomial Transformation Result

Secondly, using a polynomial transformation of degree 2, we adjust the parameter of L2 regularization and observe the results. The different regularization parameters that we apply are in the set C = {0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 1, 5, 10, 40, 60, 100}.

The result of applying different parameters on L1 regularization is provided below (Figure 16).

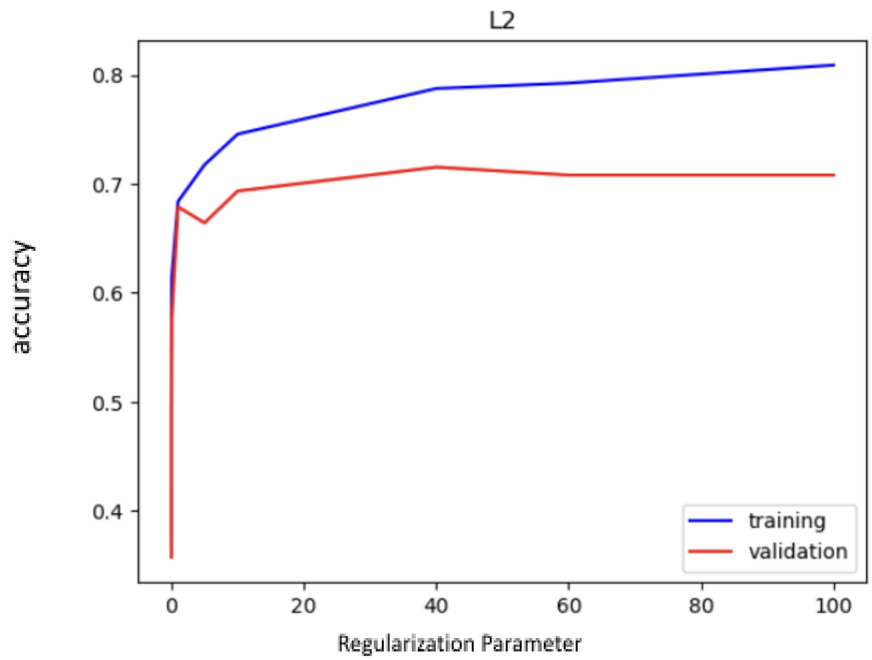

Figure 16: L2 Score

The change in feature weights when using different hyperparameters is shown below (Figure 17). Each separate line represents a weight.

Weight vs. Regularization

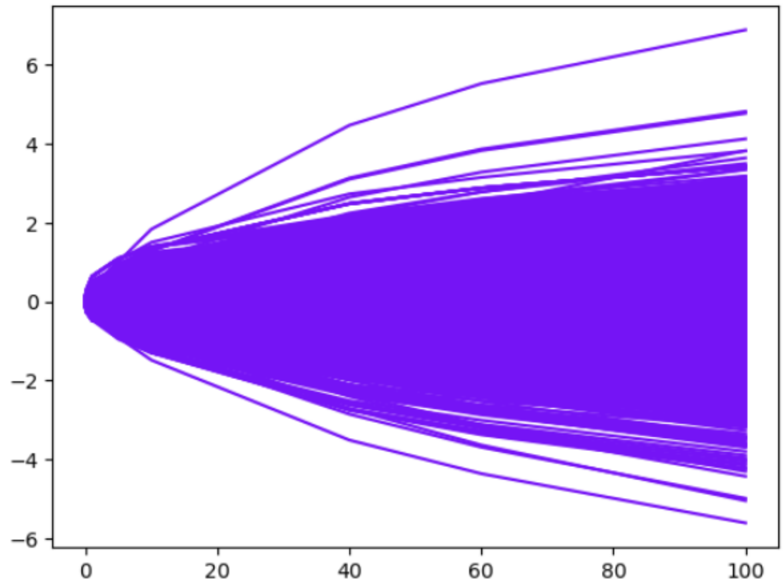

Figure 17: Feature Weight Change

We attempt to also investigate in L1 regularization; however, we never successfully compute everything we needed due to the limit of our computers. Therefore, the only regularization that we implement on logistic regression is L2.

We observe that the best setting is to have a polynomial transformation to the 7th degree without any regularization.

Then, we use this model to predict the test set and record the result scores. The results are later used to compare with other optimized models.

Linear Regression

In addition to the three classification models we use, we also use an implementation of a linear regression model. Since a regression model cannot be judged by scores like accuracy and precision, the purpose of using linear regression, in this case, is not to compare it with other models; instead, the purpose of this model is to address the issue of the dataset, that wine of quality 1, 2, 8, 9, 10 do not exist in samples. This model is implemented out of practical thinking: this model helps us to avoid situations like an excellent wine of quality 10 or 9 that we cannot identify, or a terrible wine of quality 1 or 2 is presented but we do not know it.

Firstly, we evaluate the performances of different polynomial transformations in an untuned setting. We examine the performances of transformation between degree 1 to degree 7. The result is shown below (Figure 20).

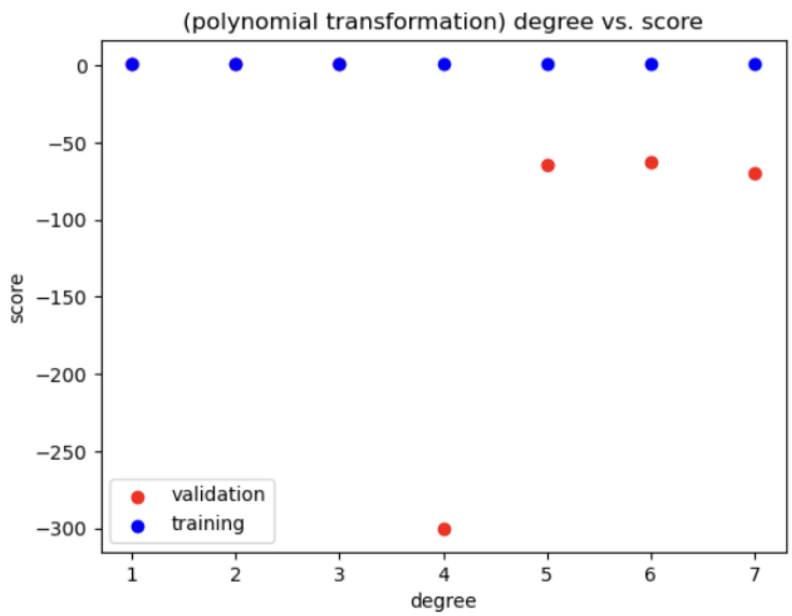

Figure 20: Polynomial Transformations Result

This graph is misleading in a way that the negative numbers have made the scale so huge that we cannot see the positive validation score, which are values in the range [0,1]. Therefore, an adjusted graph is provided below (Figure 21). We observe that degree 2 is the best-performed feature transformation. This is because it has the highest validation score, and it has the smallest difference between the validation score and training score.

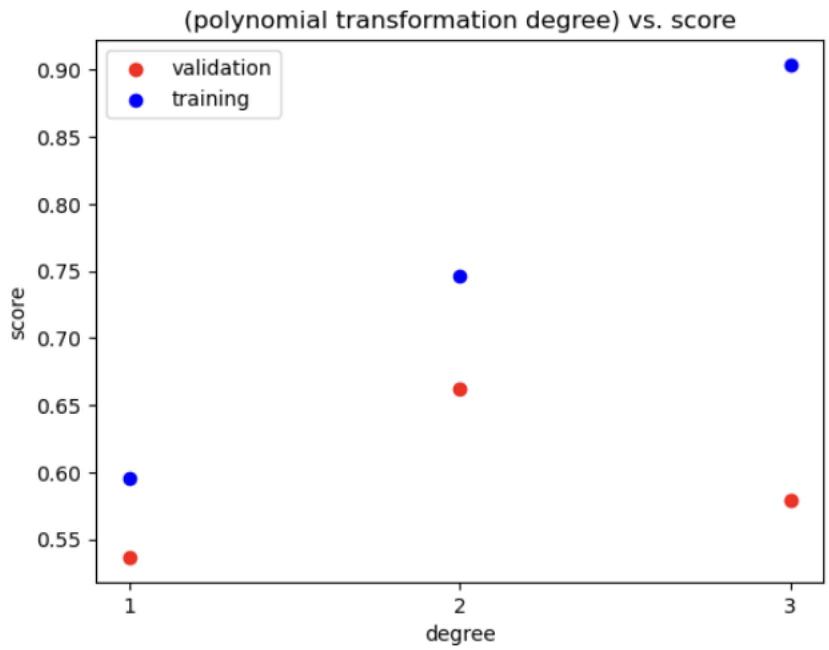

Figure 21: (Adjusted) Polynomial Transformations Result

Secondly, using a polynomial transformation of degree 2, we adjust the parameter of L1 and L2 regularization and observe the results. The different regularization parameters that we apply are in the set C = {0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 1, 5, 10, 20, 40, 60, 75, 100}.

The result of applying different parameters on L1 regularization is provided below (Figure 22).

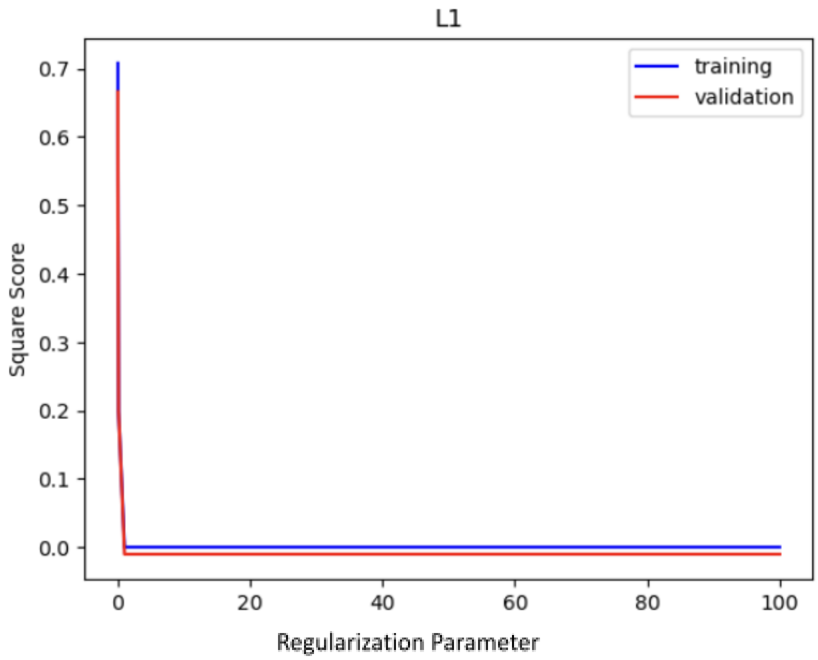

Figure 22: L1 Regularization Result

The result of applying different parameters on L2 regularization is provided below (Figure 23).

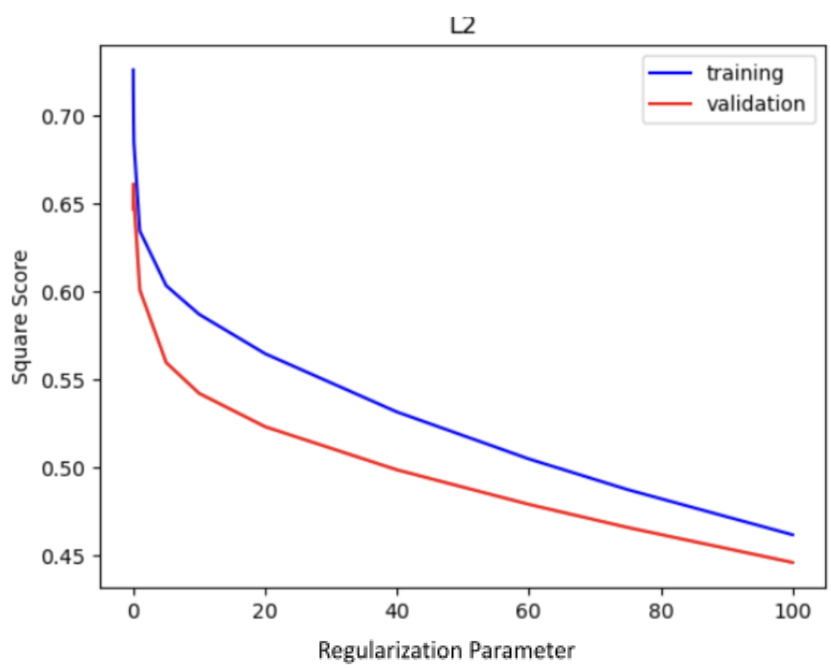

Figure 23: L2 Regularization Result

We conclude that the best-performed model is a polynomial transformation of degree 2 with an L2 regularization parameter of 0.0001.

We also record the change in feature weights when we have different regularizations and different regularization parameters.

The feature weight change when changing the parameter of L1 regularization is shown below (Figure 24). The legend is ignored because it is too long; there are 78 weights in total. It may appear that there is only a few lines; it is actually a graph of 78 lines.

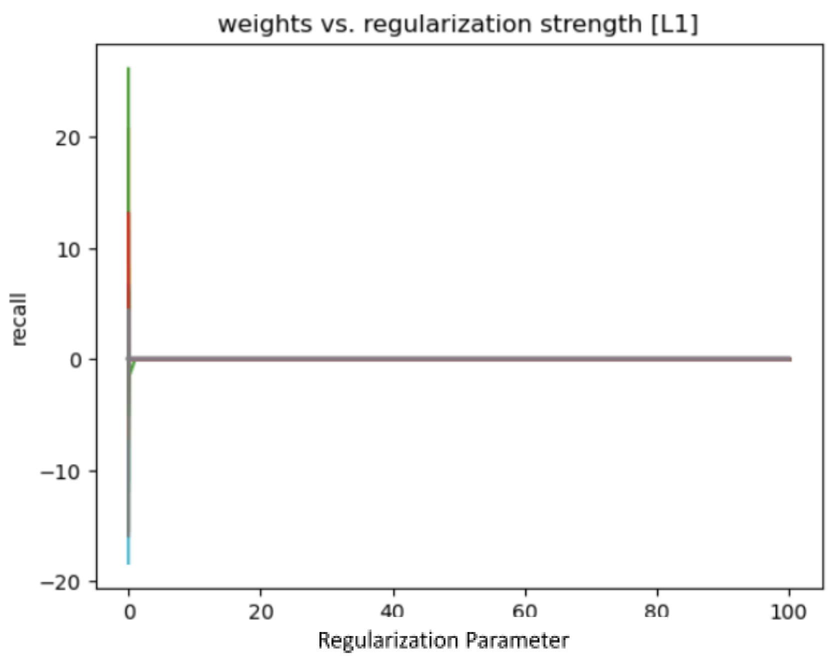

Figure 24: Weight Changes (L1)

The feature weight change when changing the parameter of L1 regularization is shown below (Figure 24). The legend is also ignored.

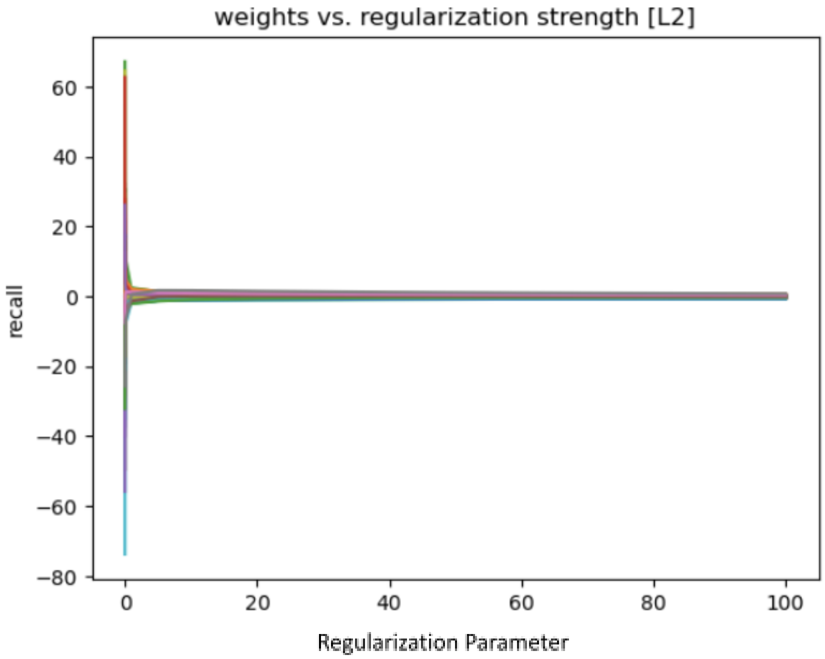

Figure 25: Weight Changes (L2)

Results:

The table below shows the data that we collect when training different models (Table 1). The yellow-color lines refer to the best validation settings and the orange-color lines refer to test settings. (Note, "NN" refers to "neural network".) The file containing all the data is also submitted separately.

Table 1: Data

2	SVM	L2	0.1	Linear	N/A	N/A	0.383810573	0.350365	N/A
3	SVM	L2	0.1	RBF	N/A	N/A	0.509911894	0.43796	N/A
4	SVM	L2	0.1	Poly Degree2	N/A	N/A	0.532489	0.48175	N/A
5	SVM	L2	0.1	Poly Degree3	N/A	N/A	0.5754405	0.50365	N/A
6	SVM	L2	0.1	Poly Degree4	N/A	N/A	0.6139868	0.58394	N/A
7	SVM	L2	0.1	Poly Degree5	N/A	N/A	0.6585903	0.63504	N/A
8	SVM	L2	0.1	Poly Degree6	N/A	N/A	0.7268722	0.68613	N/A
9	SVM	L2	0.1	Poly Degree7	N/A	N/A	0.7720264	0.74453	N/A
10	SVM	L2	0.1	Poly Degree8		N/A	0.813326	0.70803	N/A
11	SVM	L2	0.1	Poly Degree9	N/A	N/A	0.8502203	0.78102	N/A
12	SVM	L2	0.1	Poly Degree10	N/A	N/A	0.9014317	0.78102	N/A
13	SVM	L2	0.1	Poly Degree11	N/A	N/A	0.9289648	0.81752	N/A
14	SVM	L2	0.1	Poly Degree12	N/A	N/A	0.9570485	0.79562	N/A
15	SVM	L2	0.001	Poly Degree7	N/A	N/A	0.5969163	0.52555	N/A
16	SVM	L2	0.005	Poly Degree7	N/A	N/A	0.6618943	0.62044	N/A
17	SVM	L2	0.1	Poly Degree7	N/A	N/A	0.7720264	0.74453	N/A
18	SVM	L2	0.5	Poly Degree7	N/A	N/A	0.8111233	0.69343	N/A
19	SVM	L2	1	Poly Degree7	N/A	N/A	0.8386564	0.74453	N/A
20	SVM	L2	5	Poly Degree7	N/A	N/A	0.8799559	0.76642	N/A
21	SVM	L2	10	Poly Degree7	N/A	N/A	0.9069383	0.78102	N/A
22	SVM	L2	15	Poly Degree7	N/A	N/A	0.9201542	0.81752	N/A
23	SVM	L2	20	Poly Degree7	N/A	N/A	0.9234581	0.81022	N/A
24	SVM	L2	30	Poly Degree7	N/A	N/A	0.9322687	0.80292	N/A
25	SVM	L2	50	Poly Degree7	N/A	N/A	0.9443833	0.81752	N/A
26	SVM	L2	80	Poly Degree7	N/A	N/A	0.9553965	0.81752	N/A
27	SVM	L2	0.001	Poly Degree7	N/A	N/A	0.2285242	0.25547	N/A
28	SVM	L2	0.005	RBF	N/A	N/A	0.2285242	0.25547	N/A
29	SVM	L2	0.1	RBF	N/A	N/A	0.5099119	0.43796	N/A
30	SVM	L2	0.5	RBF	N/A	N/A	0.5704846	0.51095	N/A
31	SVM	L2	1	RBF	N/A	N/A	0.5963656	0.53285	N/A
32	SVM	L2	5	RBF	N/A	N/A	0.6349119	0.59854	N/A
33	SVM	L2	10	RBF	N/A	N/A	0.654185	0.62044	N/A
34	SVM	L2	15	RBF	N/A	N/A	0.6921806	0.69343	N/A
35	SVM	L2	20	RBF	N/A	N/A	0.7042952	0.68613	N/A
36	SVM	L2	30	RBF	N/A	N/A	0.7257709	0.69343	N/A
37	SVM	L2	50	RBF	N/A	N/A	0.7290749	0.69343	N/A
38	SVM	L2	80	RBF	N/A	N/A	0.7444934	0.70073	N/A
39	SVM	L2	0.001	Linear	N/A	N/A	0.2285242	0.25547	N/A
40	SVM	L2	0.005	Linear	N/A	N/A	0.2285242	0.25547	N/A
41	SVM	L2	0.1	Linear	N/A	N/A	0.3838106	0.35036	N/A
42	SVM	L2	0.5	Linear	N/A	N/A	0.5192731	0.45985	N/A
43	SVM	L2	1	Linear	N/A	N/A	0.5572687	0.46715	N/A
44	SVM	L2	5	Linear	N/A	N/A	0.597467	0.55474	N/A
45	SVM	L2	10	Linear	N/A	N/A	0.595815	0.56934	N/A
46	SVM	L2	15	Linear	N/A	N/A	0.6183921	0.63504	N/A
47	SVM	L2	20	Linear	N/A	N/A	0.6222467	0.62774	N/A
48	SVM	L2	30	Linear	N/A	N/A	0.625	0.62774	N/A
49	SVM	L2	50	Linear	N/A	N/A	0.6255507	0.62774	N/A
50	SVM	L2	80	Linear	N/A	N/A	0.6172907	0.63504	N/A
51	SVM	L2	15	Poly Degree7	N/A	N/A	N/A	N/A	0.777027
52	NN	L2	0.0001	N/A	[100]	ReLu	0.6685022	0.66423	N/A
53	NN	L2	0.0001	N/A	[100]	tanh	0.5556167	0.66423	N/A
54	NN	L2	0.0001	N/A	[100]	sigmoid	0.5011013	0.44526	N/A
55	NN	L2	0.0001	N/A	[100]	ReLu	0.7951542	0.74453	N/A
56	NN	L2	0.0005	N/A	[100]	ReLu	0.7868943	0.70803	N/A
57	NN	L2	0.001	N/A	[100]	ReLu	0.7681718	0.70073	N/A
58	NN	L2	0.005	N/A	[100]	ReLu	0.7907489	0.70803	N/A
59	NN	L2	0.01	N/A	[100]	ReLu	0.7588106	0.69343	N/A
60	NN	L2	0.05	N/A	[100]	ReLu	0.6310573	0.60584	N/A
61	NN	L2	0.1	N/A	[100]	ReLu	0.6475771	0.60584	N/A
62	NN	L2	1	N/A	[100]	ReLu	0.527533	0.54015	N/A
63	NN	L2	5	N/A	[100]	ReLu	0.2285242	0.25547	N/A
64	NN	L2	10	N/A	[100]	ReLu	0.2285242	0.25547	N/A
65	NN	L2	20	N/A	[100]	ReLu	0.215859	0.22628	N/A
66	NN	L2	75	N/A	[100]	ReLu	0.2285242	0.25547	N/A
67	NN	L2	100	N/A	[100]	ReLu	0.2285242	0.25547	N/A
68	NN	L2	0.0001	N/A	[200,40]	ReLu	0.7411894	0.70073	N/A
69	NN	L2	0.0005	N/A	[200,40]	ReLu	0.7131057	0.67153	N/A
70	NN	L2	0.001	N/A	[200,40]	ReLu	0.720815	0.68613	N/A
71	NN	L2	0.005	N/A	[200,40]	ReLu	0.7505507	0.67883	N/A
72	NN	L2	0.01	N/A	[200,40]	ReLu	0.7202643	0.67883	N/A
73	NN	L2	0.05	N/A	[200,40]	ReLu	0.6668502	0.62774	N/A

74	NN	L2	0.1	N/A	[200,40]	ReLu	0.6751101	0.67153	N/A
75	NN	L2	1	N/A	[200,40]	ReLu	0.5247797	0.51825	N/A
76	NN	L2	5	N/A	[200,40]	ReLu	0.2285242	0.25547	N/A
77	NN	L2	10	N/A	[200,40]	ReLu	0.2285242	0.25547	N/A
78	NN	L2	20	N/A	[200,40]	ReLu	0.2285242	0.25547	N/A
79	NN	L2	75	N/A	[200,40]	ReLu	0.2285242	0.25547	N/A
80	NN	L2	100	N/A	[200,40]	ReLu	0.2285242	0.25547	N/A
81	NN	L2	0.0001	N/A	[500,100,50]	ReLu	0.8023128	0.70803	N/A
82	NN	L2	0.0005	N/A	[500,100,50]	ReLu	0.8111233	0.72993	N/A
83	NN	L2	0.001	N/A	[500,100,50]	ReLu	0.751652	0.70803	N/A
84	NN	L2	0.005	N/A	[500,100,50]	ReLu	0.8182819	0.73723	N/A
85	NN	L2	0.01	N/A	[500,100,50]	ReLu	0.7692731	0.72993	N/A
86	NN	L2	0.05	N/A	[500,100,50]	ReLu	0.7890969	0.73723	N/A
87	NN	L2	0.1	N/A	[500,100,50]	ReLu	0.6657489	0.65693	N/A
88	NN	L2	1	N/A	[500,100,50]	ReLu	0.5247797	0.51095	N/A
89	NN	L2	5	N/A	[500,100,50]	ReLu	0.2285242	0.25547	N/A
90	NN	L2	10	N/A	[500,100,50]	ReLu	0.2285242	0.25547	N/A
91	NN	L2	20	N/A	[500,100,50]	ReLu	0.2285242	0.25547	N/A
92	NN	L2	75	N/A	[500,100,50]	ReLu	0.2285242	0.25547	N/A
93	NN	L2	100	N/A	[500,100,50]	ReLu	0.2285242	0.25547	N/A
94	NN	L2	0.0001	N/A	[500,200,100,30]	ReLu	0.7879956	0.71533	N/A
95	NN	L2	0.0005	N/A	[500,200,100,30]	ReLu	0.7511013	0.66423	N/A
96	NN	L2	0.001	N/A	[500,200,100,30]	ReLu	0.7912996	0.72263	N/A
97	NN	L2	0.005	N/A	[500,200,100,30]	ReLu	0.8221366	0.75182	N/A
98	NN	L2	0.01	N/A	[500,200,100,30]	ReLu	0.7935022	0.72993	N/A
99	NN	L2	0.05	N/A	[500,200,100,30]	ReLu	0.7797357	0.69343	N/A
100	NN	L2	0.1	N/A	[500,200,100,30]	ReLu	0.7599119	0.70803	N/A
101	NN	L2	1	N/A	[500,200,100,30]	ReLu	0.5401982	0.54745	N/A
102	NN	L2	5	N/A	[500,200,100,30]	ReLu	0.2285242	0.25547	N/A
103	NN	L2	10	N/A	[500,200,100,30]	ReLu	0.2285242	0.25547	N/A
104	NN	L2	20	N/A	[500,200,100,30]	ReLu	0.2285242	0.25547	N/A
105	NN	L2	75	N/A	[500,200,100,30]	ReLu	0.2285242	0.25547	N/A
106	NN	L2	100	N/A	[500,200,100,30]	ReLu	0.2285242	0.25547	N/A
107	NN	L2	0.0001	N/A	[500,250,125,70,35]	ReLu	0.7951542	0.74453	N/A
108	NN	L2	0.0005	N/A	[500,250,125,70,35]	ReLu	0.7868943	0.70803	N/A
109	NN	L2	0.001	N/A	[500,250,125,70,35]	ReLu	0.7681718	0.70073	N/A
110	NN	L2	0.005	N/A	[500,250,125,70,35]	ReLu	0.7907489	0.70803	N/A
111	NN	L2	0.01	N/A	[500,250,125,70,35]	ReLu	0.7588106	0.69343	N/A
112	NN	L2	0.05	N/A	[500,250,125,70,35]	ReLu	0.6310573	0.60584	N/A
113	NN	L2	0.1	N/A	[500,250,125,70,35]	ReLu	0.6475771	0.60584	N/A
114	NN	L2	1	N/A	[500,250,125,70,35]	ReLu	0.527533	0.54015	N/A
115	NN	L2	5	N/A	[500,250,125,70,35]	ReLu	0.2285242	0.25547	N/A
116	NN	L2	10	N/A	[500,250,125,70,35]	ReLu	0.2285242	0.25547	N/A
117	NN	L2	20	N/A	[500,250,125,70,35]	ReLu	0.215859	0.22628	N/A
118	NN	L2	75	N/A	[500,250,125,70,35]	ReLu	0.2285242	0.25547	N/A
119	NN	L2	100	N/A	[500,250,125,70,35]	ReLu	0.2285242	0.25547	N/A
120	NN	L2	0.0001	N/A	[700,500,300,200,150,100,50]	ReLu	0.7885463	0.71533	N/A
121	NN	L2	0.0005	N/A	[700,500,300,200,150,100,50]	ReLu	0.7725771	0.68613	N/A
122	NN	L2	0.001	N/A	[700,500,300,200,150,100,50]	ReLu	0.7571586	0.74453	N/A
123	NN	L2	0.005	N/A	[700,500,300,200,150,100,50]	ReLu	0.8177313	0.72993	N/A
124	NN	L2	0.01	N/A	[700,500,300,200,150,100,50]	ReLu	0.7329295	0.67883	N/A
125	NN	L2	0.05	N/A	[700,500,300,200,150,100,50]	ReLu	0.7764317	0.71533	N/A
126	NN	L2	0.1	N/A	[700,500,300,200,150,100,50]	ReLu	0.7048458	0.64964	N/A
127	NN	L2	1	N/A	[700,500,300,200,150,100,50]	ReLu	0.2285242	0.25547	N/A
128	NN	L2	5	N/A	[700,500,300,200,150,100,50]	ReLu	0.2285242	0.25547	N/A
129	NN	L2	10	N/A	[700,500,300,200,150,100,50]	ReLu	0.2285242	0.25547	N/A
130	NN	L2	20	N/A	[700,500,300,200,150,100,50]	ReLu	0.2285242	0.25547	N/A
131	NN	L2	75	N/A	[700,500,300,200,150,100,50]	ReLu	0.2285242	0.25547	N/A
132	NN	L2	100	N/A	[700,500,300,200,150,100,50]	ReLu	0.2285242	0.25547	N/A
133	NN	L2	0.005	N/A	[500,200,100,30]	ReLu	N/A	N/A	0.783784
134	Linear	None	N/A	Linear	N/A	N/A	0.5953987	0.53652	N/A
135	Linear	None	N/A	Poly Degree 2	N/A	N/A	0.7462198	0.66211	N/A
136	Linear	None	N/A	Poly Degree 3	N/A	N/A	0.9037507	0.57906	N/A
137	Linear	None	N/A	Poly Degree 4	N/A	N/A	1	-300.2	N/A
138	Linear	None	N/A	Poly Degree 5	N/A	N/A	1	-64.796	N/A
139	Linear	None	N/A	Poly Degree 6	N/A	N/A	1	-62.76	N/A
140	Linear	None	N/A	Poly Degree 7	N/A	N/A	1	-70.193	N/A
141	Linear	L2	0.0001	Poly Degree 2	N/A	N/A	0.7261784	0.72466	N/A
142	Linear	L2	0.0005	Poly Degree 2	N/A	N/A	0.7246566	0.65054	N/A
143	Linear	L2	0.001	Poly Degree 2	N/A	N/A	0.7238387	0.65268	N/A
144	Linear	L2	0.005	Poly Degree 2	N/A	N/A	0.7194538	0.65834	N/A
145	Linear	L2	0.01	Poly Degree 2	N/A	N/A	0.7155499	0.66126	N/A

146	Linear	L2	0.05	Poly Degree 2	N/A	N/A	0.69764589	0.660671	N/A
147	Linear	L2	0.1	Poly Degree 2	N/A	N/A	0.68504289	0.65292	N/A
148	Linear	L2	1	Poly Degree 2	N/A	N/A	0.63451229	0.600976	N/A
149	Linear	L2	5	Poly Degree 2	N/A	N/A	0.60340418	0.559652	N/A
150	Linear	L2	10	Poly Degree 2	N/A	N/A	0.5871081	0.542129	N/A
151	Linear	L2	20	Poly Degree 2	N/A	N/A	0.56471616	0.523167	N/A
152	Linear	L2	40	Poly Degree 2	N/A	N/A	0.53154318	0.498608	N/A
153	Linear	L2	60	Poly Degree 2	N/A	N/A	0.5048288	0.478979	N/A
154	Linear	L2	75	Poly Degree 2	N/A	N/A	0.4873404	0.465813	N/A
155	Linear	L2	100	Poly Degree 2	N/A	N/A	0.46167998	0.445912	N/A
156	Linear	L1	0.0001	Poly Degree 2	N/A	N/A	0.70812658	0.66605	N/A
157	Linear	L1	0.0005	Poly Degree 2	N/A	N/A	0.65798471	0.638084	N/A
158	Linear	L1	0.001	Poly Degree 2	N/A	N/A	0.61902312	0.58405	N/A
159	Linear	L1	0.005	Poly Degree 2	N/A	N/A	0.57702181	0.52889	N/A
160	Linear	L1	0.01	Poly Degree 2	N/A	N/A	0.54102713	0.484458	N/A
161	Linear	L1	0.05	Poly Degree 2	N/A	N/A	0.40816449	0.39118814	N/A
162	Linear	L1	0.1	Poly Degree 2	N/A	N/A	0.19822068	0.19835448	N/A
163	Linear	L1	1	Poly Degree 2	N/A	N/A	0	-0.01033137	N/A
164	Linear	L1	5	Poly Degree 2	N/A	N/A	0	-0.01033137	N/A
165	Linear	L1	10	Poly Degree 2	N/A	N/A	0	-0.01033137	N/A
166	Linear	L1	20	Poly Degree 2	N/A	N/A	0	-0.01033137	N/A
167	Linear	L1	40	Poly Degree 2	N/A	N/A	0	-0.01033137	N/A
168	Linear	L1	60	Poly Degree 2	N/A	N/A	0	-0.01033137	N/A
169	Linear	L1	75	Poly Degree 2	N/A	N/A	0	-0.01033137	N/A
170	Linear	L1	100	Poly Degree 2	N/A	N/A	0	-0.01033137	N/A
171	Linear	L2	0.0001	Poly Degree 2	N/A	N/A	N/A	N/A	0.6260887
172	Logistic	L2	1	Linear	N/A	N/A	0.58755507	0.569343	N/A
173	Logistic	L2	1	Poly Degree 2	N/A	N/A	0.75660793	0.642336	N/A
174	Logistic	L2	1	Poly Degree 3	N/A	N/A	0.81828194	0.678832	N/A
175	Logistic	L2	1	Poly Degree 4	N/A	N/A	0.85737885	0.722628	N/A
176	Logistic	L2	1	Poly Degree 5	N/A	N/A	0.86178414	0.737226	N/A
177	Logistic	L2	1	Poly Degree 6	N/A	N/A	0.87555066	0.744526	N/A
178	Logistic	L2	1	Poly Degree 7	N/A	N/A	0.89482379	0.751825	N/A
179	Logistic	L2	0.0001	Poly Degree 7	N/A	N/A	0.36839207	0.357664	N/A
180	Logistic	L2	0.0005	Poly Degree 7	N/A	N/A	0.42015419	0.372263	N/A
181	Logistic	L2	0.001	Poly Degree 7	N/A	N/A	0.4746696	0.430657	N/A
182	Logistic	L2	0.005	Poly Degree 7	N/A	N/A	0.54570485	0.49635	N/A
183	Logistic	L2	0.01	Poly Degree 7	N/A	N/A	0.5589207	0.49635	N/A
184	Logistic	L2	0.05	Poly Degree 7	N/A	N/A	0.61343612	0.576642	N/A
185	Logistic	L2	0.1	Poly Degree 7	N/A	N/A	0.61839207	0.583942	N/A
186	Logistic	L2	1	Poly Degree 7	N/A	N/A	0.6839207	0.678832	N/A
187	Logistic	L2	5	Poly Degree 7	N/A	N/A	0.71751101	0.664234	N/A
188	Logistic	L2	10	Poly Degree 7	N/A	N/A	0.74559471	0.693431	N/A
189	Logistic	L2	40	Poly Degree 7	N/A	N/A	0.78744493	0.715328	N/A
190	Logistic	L2	60	Poly Degree 7	N/A	N/A	0.79240088	0.708029	N/A
191	Logistic	L2	75	Poly Degree 7	N/A	N/A	0.8089207	0.708029	N/A
192	Logistic	L2	100	Poly Degree 7	N/A	N/A	0.8089207	0.708029	N/A
193	Logistic	L2	40	Poly Degree 7	N/A	N/A	N/A	N/A	0.8243243
194									

Conclusions:

The optimized setting for the SVM is using L2 regularization with regularization parameter = 15 under a polynomial kernel to the 7th degree. The optimized setting for the neural network is a network with 4 hidden layers of structure (500, 200, 100, 30) with a ReLu activation using L2 regularization with a regularization parameter = 0.005. The optimized setting for the logistic regression is a polynomial feature transformation to the 7th degree without any regularization.

The test score of the optimized SVM model is 0.777027027027027; the test score of the optimized neural network model is 0.7837837837837838; the test score of the optimized logistic regression model is 0.8243243243243243; the test score of the optimized linear regression model is 0.6260886676461701.

Since linear regression is a regression problem and is measured in different metrics, we will not compare it with the classification model. Therefore, the model that performs the best is logistic regression and the model that performs the worst is SVM. In fact, the performance of the logistic regression can be further improved because, after 10000 iterations on the test set, we still have not converged yet. We decide not to continue the iterations because we have made our point with the current result.

We observe that the logistic regression and SVM models have huge gaps between training and validation scores when using polynomial feature transformation; the training scores are significantly higher than the validation scores. From the observation, we conclude that polynomial feature transformation is potentially causing overfitting when applied to logistic regression and SVM.

In the case of SVM, among the three kernels that we try, RBF, linear, and polynomial, none of those performed ideally. RBF and linear kernel results in low training and validation score; in another word, high bias and variance, which means underfitting. Even though polynomial kernel overfits, it is already the best-performed choice in terms of validation score. Since none of the feature transformations actually work, we deduce that there are two possible ways to improve the SVM model: first, come up with a different kernel; decrease the noise from the original dataset.

In the case of logistic regression, we are not 100% confident that we are dealing with overfitting despite what the result suggests. We do not want to conclude that the model overfits because, in most settings, the logistic regression model fails to converge (we have given as many iterations as our computers allow). We do not know if the logistic regression models will still overfit when they are fully trained to converge. There is even a possibility that the optimized model underfits because we do not know whether the performance of the model would increase if we try polynomial transformation of higher degrees (we did not try anything above degree 7 due to the limits of our computers).

We observe an interesting phenomenon when we analyze the results of different layers of neural networks: all networks perform better when the regularization parameters are smaller—even the training scores are higher when the regularization parameters are small. This means that, in general, on this dataset neural network models perform better when the regularization strength is great. Less regularization strength not only results in higher variance but also higher bias. Usually, less regularization strength causes overfitting; however, in our case, less regularization strength causes underfits. We do not fully understand the phenomenon. We only have one theory: none of the activation functions suits the dataset.

Among all four optimized models, none of them uses a polynomial transformation of less than degree 2 or less than 2 hidden layers, which suggests that the dataset is not linearly separable. However, in different models, the dataset reflects different levels of complexity. Therefore, we believe that finding a universal complexity degree—a complexity degree in which all the models would perform very well, improves the model performances; this is because the universal complexity degree will give us a good approximation of the true picture of the dataset.

Another thing that we could try is to find a better computer and reevaluate the performances of different settings applied to logistic regression. Logistic regression is our best-performed model so far—even when it fails to converge in most settings, so we think fully training the logistic regression model would significantly improve our performance.