Unit Project 11: College Application Calculator

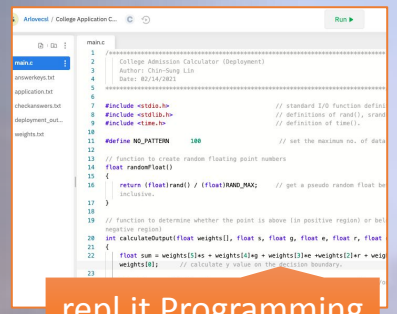
Chin-Sung Lin

Eleanor Roosevelt High School

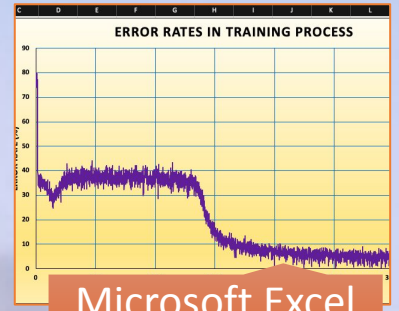
Aims of College Application Calculator Project

- Apply Perceptron to Classify Multi-Dimensional Binary Separable Data
- Understand the Steps to Design an AI system
- Use AI Technique to Solve a Complex Real-World Problem
- Develop Insights of the Power & Limitations of AI
- Explore Potential Real-Life Applications of AI
- Recognize Possible Social Impacts of AI Technology

Tools & Resources



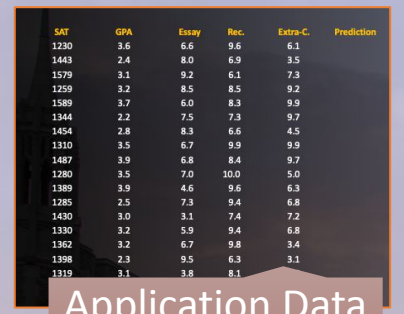
repl.it Programming Environment



Microsoft Excel Spreadsheet

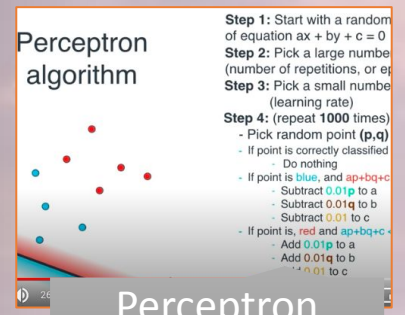
SAT	GPA	Essay	Rec.	Extra-C.	Decision
1579	2.4	9.3	8.7	3.6	R
1312	2.2	5.2	7.4	3.3	R
1372	3.8	6.6	7.0	5.6	R
1517	2.2	9.0	8.7	4.5	R
1396	3.3	3.1	8.6	3.5	R
1257	3.2	5.1	7.7	7.6	R
1327	2.7	9.4	6.9	2.4	R
1246	3.0	3.9	7.7	8.8	R
1314	3.7	8.0	9.2	2.9	R
1266	3.2	4.5	6.1	4.1	R
1409	3.0	9.4	8.1	3.3	R
1240	3.4	8.3	7.6	2.8	R
1379	3.5	6.4	7.4	9.6	A
1411	3.0	6.1	8.6	9.3	R
1317	3.9	8.4	6.1		A
1551					

Training & Testing Data Set



SAT	GPA	Essay	Rec.	Extra-C.	Prediction
1230	3.6	6.6	9.6	6.1	
1443	2.4	8.0	6.9	3.5	
1579	3.1	9.2	6.1	7.3	
1259	3.2	8.5	8.5	9.2	
1589	3.7	6.0	8.3	9.9	
1344	2.2	7.5	7.3	9.7	
1454	2.8	8.3	6.6	4.5	
1310	3.5	6.7	9.9	9.9	
1487	3.9	6.8	8.4	9.7	
1280	3.5	7.0	10.0	5.0	
1389	3.9	4.6	9.6	6.3	
1285	2.5	7.3	9.4	6.8	
1430	3.0	3.1	7.4	7.2	
1330	3.2	5.9	9.4	6.8	
1362	3.2	6.7	9.8	3.4	
1398	2.3	9.5	6.3	3.1	
1319	3.1	3.8	8.1		

Application Data Set



Perceptron Video Tutorial

Project Goals

College Application Calculator

Project Description:

- Based on a given dataset of college application & admission data of a highly competitive college from former students, develop a Perceptron model to predict the admission results of current students.

Given Dataset of Former Students:

- 1000 pieces of application & admission data.
- include students' SAT (1600 scale, >= 1200), GPA (4.0 scale), Essay (0.0-10.0), Rec. Letters (0.0-10.0), Extra-Curricular Activities (0.0-10.0) & Decision (A- acceptance, R- rejection).

Given Application Dataset of Current Students:

- 100 pieces of application data.
- include students' SAT (1600 scale, >= 1200), GPA (4.0 scale), Essay (0.0-10.0), Rec. Letters (0.0-10.0) & Extra-Curricular Activities (0.0-10.0).

Project Applications

High School Student Applicants:

- Help student evaluate their application qualifications before starting formal application.

High School College Counselors:

- Provide data-driven professional advises to student applicants.

College Admission Officers:

- Automate and accelerate the decision-making process & cut operating cost.

Former Student Application Data Set

SAT	GPA	Essay	Rec.		Extra-C.	Decision
1579	2.4	9.3	8.7	3.6	R	
1312	2.2	5.2	7.4	3.3	R	
1372	3.8	6.6	7.0	5.6	R	
1517	2.2	9.0	8.7	4.5	R	
1396	3.3	3.1	8.6	3.5	R	
1257	3.2	5.1	7.7	7.6	R	
1327	2.7	9.4	6.9	2.4	R	
1246	3.0	3.9	7.7	8.8	R	
1314	3.7	8.0	9.2	2.9	R	
1266	3.2	4.5	6.1	4.1	R	
1409	3.0	9.4	8.1	3.3	R	
1240	3.4	8.3	7.6	2.8	R	
1379	3.5	6.4	7.4	9.6	A	
1411	3.0	6.1	8.6	9.3	R	
1317	3.9	8.4	6.1	7.9	A	
1551	2.4	5.3	6.4	3.7	R	
1432	3.2	8.1	8.0	3.2	R	
1553	3.2	8.8	8.5	2.1	R	
1573	2.2	8.2	7.6	5.6	R	
1473	3.9	9.5	6.5	6.8	A	
1533	2.9	6.8	8.4	5.5	R	

SAT	GPA	Essay	Rec.		Extra-C.	Decision
1309	3.0	7.3	8.4	6.6	R	
1531	2.4	4.3	8.2	7.4	R	
1335	2.9	4.9	6.7	2.6	R	
1313	2.2	4.1	6.1	6.1	R	
1441	3.4	6.2	8.1	8.7	R	
1219	2.7	4.8	8.4	9.7	R	
1480	3.7	6.1	7.3	5.7	R	
1208	2.3	7.5	6.8	7.7	R	
1328	3.1	4.0	8.4	7.7	R	
1286	3.8	8.7	7.5	9.2	A	
1332	4.0	7.2	9.1	6.0	A	
1374	3.7	9.1	8.8	5.4	R	
1534	2.8	5.0	7.1	7.8	R	
1498	2.6	9.1	7.6	6.1	R	
1433	3.4	3.5	8.9	4.5	R	
1515	3.9	4.3	8.4	4.6	R	
1233	3.9	5.1	8.7	7.8	A	
1521	2.2	6.9	8.7	8.6	R	
1591	3.0	4.6	7.0	8.3	R	
1578	2.0	3.7	9.3	5.2	R	
1446	2.8	3.7	8.7	3.1	R	

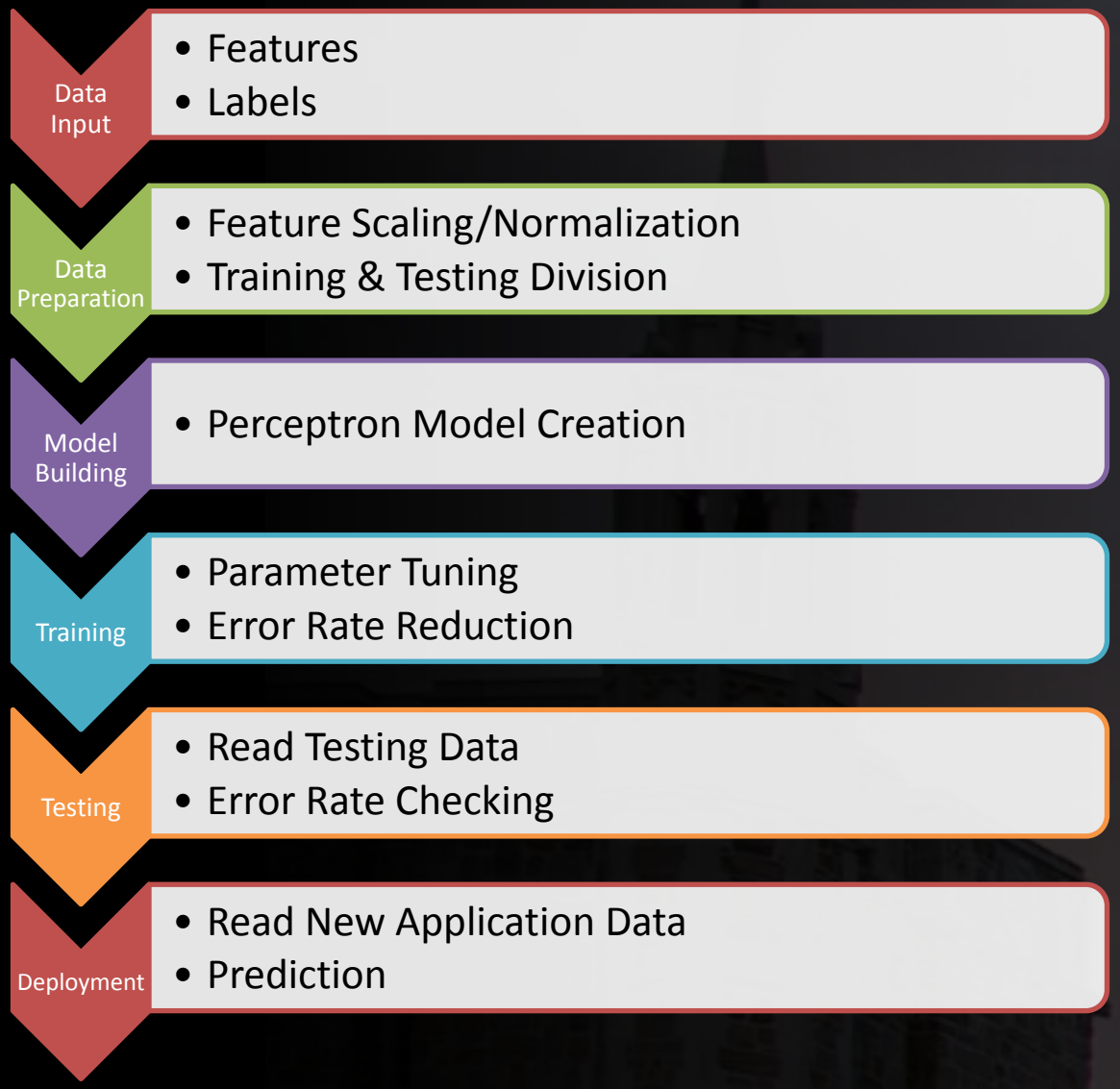
Current Student Application Data Set

SAT	GPA	Essay	Rec.	Extra-C.	Prediction
1555	3.6	4.7	8.4	5.1	
1384	3.2	5.8	7.3	7.5	
1582	3.3	4.8	9.8	7.7	
1455	2.3	8.3	8.2	7.6	
1309	3.6	8.7	9.7	8.9	
1522	3.8	3.6	6.5	4.1	
1272	2.0	3.5	7.7	6.8	
1381	3.8	4.5	9.4	3.8	
1558	3.6	9.0	6.6	8.2	
1427	3.6	9.4	7.3	4.7	
1446	3.2	3.8	7.7	6.1	
1487	3.7	8.5	8.3	5.9	
1207	2.4	8.0	6.2	8.0	
1300	2.8	9.4	7.6	9.3	
1403	2.1	5.6	7.4	2.8	
1567	3.1	8.8	8.9	2.7	
1322	2.4	7.3	9.9	3.9	
1330	3.4	3.6	6.4	4.2	
1430	2.3	6.3	7.1	3.4	
1290	3.1	7.2	6.6	9.4	
1402	3.3	9.8	9.5	9.9	

SAT	GPA	Essay	Rec.	Extra-C.	Prediction
1230	3.6	6.6	9.6	6.1	
1443	2.4	8.0	6.9	3.5	
1579	3.1	9.2	6.1	7.3	
1259	3.2	8.5	8.5	9.2	
1589	3.7	6.0	8.3	9.9	
1344	2.2	7.5	7.3	9.7	
1454	2.8	8.3	6.6	4.5	
1310	3.5	6.7	9.9	9.9	
1487	3.9	6.8	8.4	9.7	
1280	3.5	7.0	10.0	5.0	
1389	3.9	4.6	9.6	6.3	
1285	2.5	7.3	9.4	6.8	
1430	3.0	3.1	7.4	7.2	
1330	3.2	5.9	9.4	6.8	
1362	3.2	6.7	9.8	3.4	
1398	2.3	9.5	6.3	3.1	
1319	3.1	3.8	8.1	5.6	
1459	3.5	8.0	7.1	6.7	
1328	3.7	3.6	7.3	3.4	
1495	3.3	8.6	6.6	6.1	
1356	3.1	3.7	9.7	6.1	

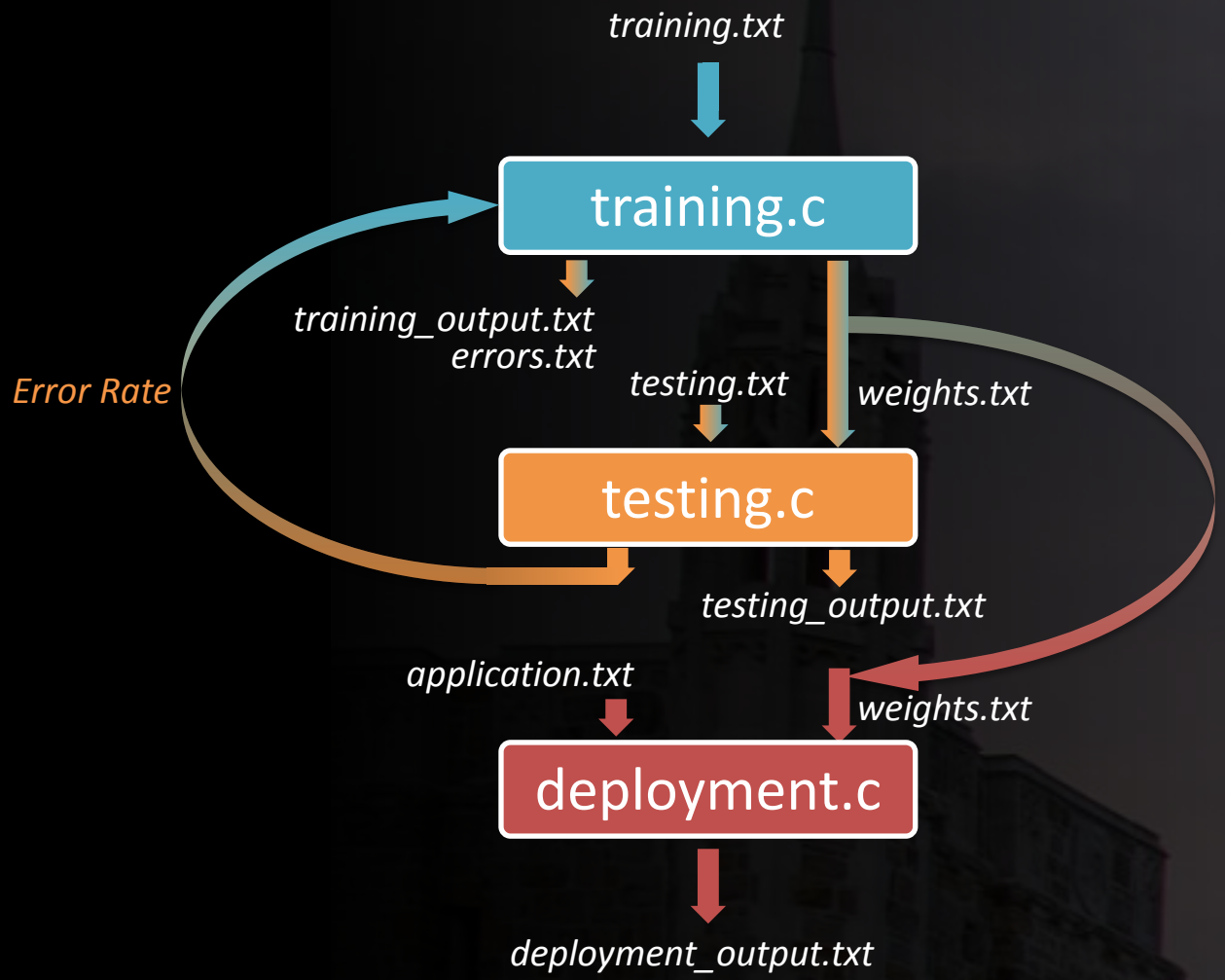
Project Flow

Machine Learning Classification Process



Project Deliverables

Training ⟹ Testing ⟹ Deployment



Training Output: Decision Boundary

(training_output.txt)

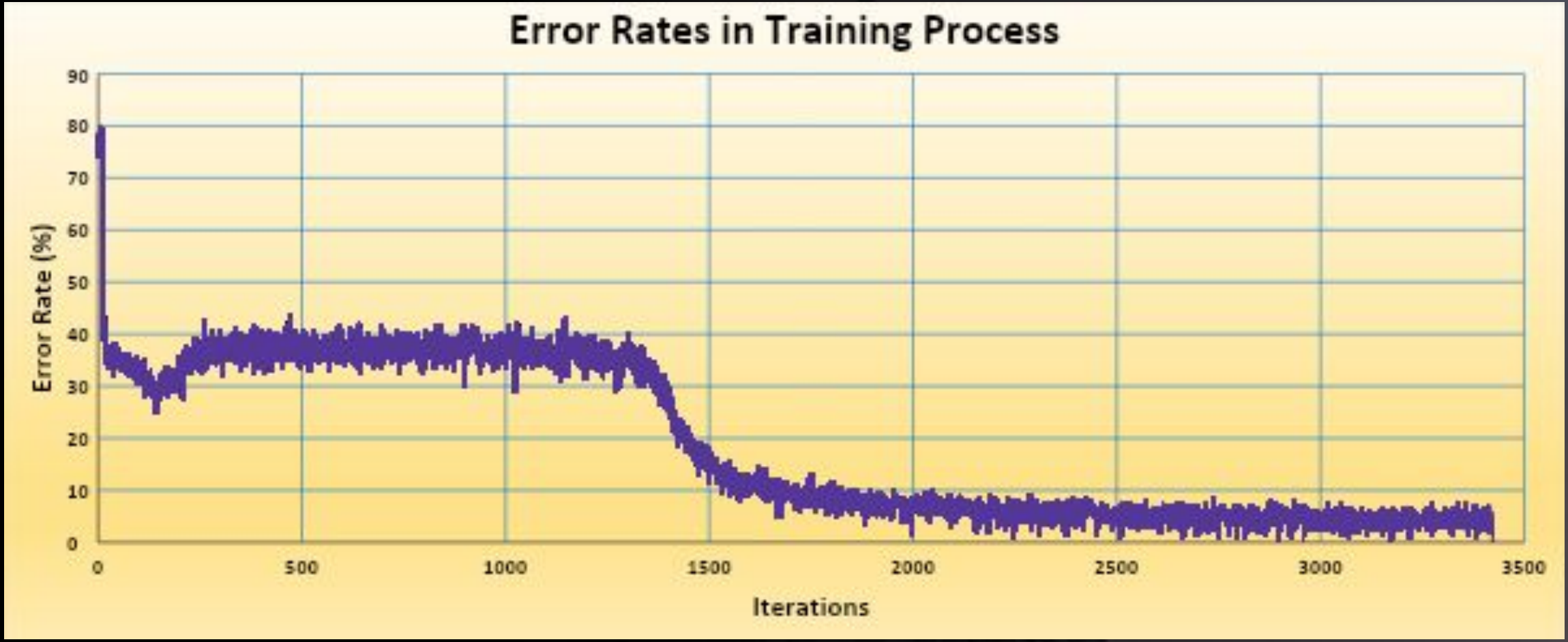
```
Initial Equation :   9.0 * s + 2.8 * g + 6.2 * e + 5.9 * r + 4.6 * c + 5.5 = 0

Iteration   0- Equation: 8.769*s + 2.354*g + 5.828*e + 5.451*r + 4.250*c + 5.424 = 0     ERR = 77.38 %
Iteration   1- Equation: 8.491*s + 1.918*g + 5.424*e + 4.955*r + 3.896*c + 5.362 = 0     ERR = 78.25 %
Iteration   2- Equation: 8.242*s + 1.507*g + 5.039*e + 4.482*r + 3.568*c + 5.302 = 0     ERR = 73.88 %
Iteration   3- Equation: 7.987*s + 1.060*g + 4.625*e + 3.975*r + 3.203*c + 5.238 = 0     ERR = 79.88 %
Iteration   4- Equation: 7.731*s + 0.633*g + 4.230*e + 3.489*r + 2.863*c + 5.177 = 0     ERR = 76.25 %
Iteration   5- Equation: 7.475*s + 0.207*g + 3.839*e + 3.004*r + 2.529*c + 5.116 = 0     ERR = 76.25 %
Iteration   6- Equation: 7.194*s + -0.227*g + 3.426*e + 2.498*r + 2.167*c + 5.053 = 0    ERR = 79.50 %
Iteration   7- Equation: 6.969*s + -0.637*g + 3.042*e + 2.030*r + 1.830*c + 4.993 = 0    ERR = 74.25 %
Iteration   8- Equation: 6.692*s + -1.079*g + 2.625*e + 1.523*r + 1.471*c + 4.929 = 0    ERR = 79.62 %
Iteration   9- Equation: 6.435*s + -1.512*g + 2.224*e + 1.025*r + 1.117*c + 4.867 = 0    ERR = 77.62 %
Iteration  10- Equation: 6.183*s + -1.940*g + 1.825*e + 0.537*r + 0.772*c + 4.806 = 0    ERR = 76.38 %
Iteration  11- Equation: 5.921*s + -2.368*g + 1.420*e + 0.042*r + 0.425*c + 4.744 = 0    ERR = 77.25 %
Iteration  12- Equation: 5.663*s + -2.730*g + 1.075*e + -0.377*r + 0.135*c + 4.691 = 0   ERR = 67.25 %
Iteration  13- Equation: 5.425*s + -2.998*g + 0.802*e + -0.702*r + -0.081*c + 4.650 = 0  ERR = 54.38 %
Iteration  14- Equation: 5.232*s + -3.171*g + 0.615*e + -0.925*r + -0.219*c + 4.622 = 0  ERR = 44.50 %
Iteration  15- Equation: 5.079*s + -3.283*g + 0.486*e + -1.076*r + -0.307*c + 4.602 = 0  ERR = 39.88 %
[redacted]
Iteration 3419- Equation: 0.008*s + 0.025*g + 0.003*e + 0.003*r + 0.007*c + -0.366 = 0   ERR = 4.62 %
Iteration 3420- Equation: 0.009*s + 0.025*g + 0.005*e + 0.002*r + 0.008*c + -0.366 = 0   ERR = 5.75 %
Iteration 3421- Equation: 0.008*s + 0.025*g + 0.004*e + 0.003*r + 0.006*c + -0.366 = 0   ERR = 2.12 %
Iteration 3422- Equation: 0.008*s + 0.025*g + 0.003*e + 0.004*r + 0.008*c + -0.366 = 0   ERR = 4.75 %
Iteration 3423- Equation: 0.008*s + 0.025*g + 0.003*e + 0.004*r + 0.008*c + -0.366 = 0   ERR = 0.00 %

Final Equation: 2.445*s + 7.461*g + 1.000*e + 1.047*r + 2.300*c + -107.146 = 0   ERR = 0.00 %
```

Training Process: Error Rate Reduction

(training_errors.xlsx)



Testing Output: Errors & Error Rate

(testing_output.txt)

```
Decision Boundary : 2.445*s + 7.461*g + 1.000*e + 1.047*r + 2.300*c + -107.146 = 0

000) SAT: 1490  GPA: 3.0  Essay: 8.1  Rec: 6.7  ExtraC: 7.4  Decision: R Class: R
001) SAT: 1271  GPA: 2.1  Essay: 7.0  Rec: 9.5  ExtraC: 8.2  Decision: R Class: R
002) SAT: 1333  GPA: 2.6  Essay: 7.3  Rec: 9.0  ExtraC: 9.5  Decision: R Class: R
003) SAT: 1280  GPA: 4.0  Essay: 7.2  Rec: 6.6  ExtraC: 6.1  Decision: R Class: A *** ERROR
004) SAT: 1462  GPA: 3.2  Essay: 8.1  Rec: 8.3  ExtraC: 9.5  Decision: A Class: A
005) SAT: 1581  GPA: 3.6  Essay: 4.1  Rec: 8.5  ExtraC: 6.0  Decision: A Class: A
006) SAT: 1477  GPA: 2.7  Essay: 3.1  Rec: 7.7  ExtraC: 6.3  Decision: R Class: R
007) SAT: 1472  GPA: 3.2  Essay: 7.0  Rec: 7.0  ExtraC: 5.8  Decision: R Class: R
008) SAT: 1343  GPA: 3.2  Essay: 8.4  Rec: 9.9  ExtraC: 4.7  Decision: R Class: R
[redacted]
194) SAT: 1521  GPA: 2.9  Essay: 7.5  Rec: 7.4  ExtraC: 5.6  Decision: R Class: R
195) SAT: 1248  GPA: 2.8  Essay: 3.8  Rec: 8.2  ExtraC: 3.6  Decision: R Class: R
196) SAT: 1366  GPA: 3.6  Essay: 6.1  Rec: 7.0  ExtraC: 2.3  Decision: R Class: R
197) SAT: 1221  GPA: 2.7  Essay: 9.5  Rec: 7.8  ExtraC: 9.0  Decision: R Class: R
198) SAT: 1244  GPA: 2.6  Essay: 5.5  Rec: 6.9  ExtraC: 5.7  Decision: R Class: R
199) SAT: 1322  GPA: 2.3  Essay: 3.9  Rec: 6.5  ExtraC: 5.5  Decision: R Class: R

Error Rate = 2.50 %
```

Deployment Output: Predictions

(deployment_output.txt)

```
000) SAT: 1239  GPA: 3.9  Essay: 9.1  Rec: 9.0  ExtraC: 4.3   Prediction: R
001) SAT: 1330  GPA: 3.7  Essay: 7.8  Rec: 7.8  ExtraC: 5.4   Prediction: R
002) SAT: 1552  GPA: 3.7  Essay: 4.6  Rec: 7.3  ExtraC: 2.9   Prediction: A
003) SAT: 1306  GPA: 2.8  Essay: 6.2  Rec: 6.8  ExtraC: 8.6   Prediction: R
004) SAT: 1330  GPA: 2.6  Essay: 4.0  Rec: 8.7  ExtraC: 6.3   Prediction: R
005) SAT: 1439  GPA: 4.0  Essay: 7.7  Rec: 8.9  ExtraC: 3.0   Prediction: A
006) SAT: 1243  GPA: 3.6  Essay: 3.4  Rec: 9.9  ExtraC: 6.5   Prediction: R
007) SAT: 1338  GPA: 2.6  Essay: 6.0  Rec: 6.1  ExtraC: 8.0   Prediction: R
008) SAT: 1543  GPA: 3.8  Essay: 7.3  Rec: 6.3  ExtraC: 3.8   Prediction: A
009) SAT: 1489  GPA: 2.7  Essay: 7.2  Rec: 6.7  ExtraC: 6.4   Prediction: R
010) SAT: 1371  GPA: 3.0  Essay: 9.0  Rec: 8.3  ExtraC: 3.5   Prediction: R
011) SAT: 1355  GPA: 2.3  Essay: 4.2  Rec: 6.2  ExtraC: 9.1   Prediction: R
012) SAT: 1318  GPA: 2.3  Essay: 8.0  Rec: 7.4  ExtraC: 3.2   Prediction: R
```

```
095) SAT: 1245  GPA: 3.3  Essay: 7.9  Rec: 9.7  ExtraC: 3.3   Prediction: R
096) SAT: 1498  GPA: 2.3  Essay: 3.4  Rec: 8.8  ExtraC: 2.4   Prediction: R
097) SAT: 1239  GPA: 3.9  Essay: 6.6  Rec: 7.3  ExtraC: 5.3   Prediction: R
098) SAT: 1236  GPA: 3.6  Essay: 7.3  Rec: 6.2  ExtraC: 8.2   Prediction: R
099) SAT: 1492  GPA: 2.1  Essay: 7.5  Rec: 8.6  ExtraC: 5.7   Prediction: R
```

Feature Scaling / Normalization

Former Student Application Data Set

Features

SAT	GPA	Essay	Rec.		Extra-C.	Decision
1579	2.4	9.3	8.7	3.6	R	
1312	2.2	5.2	7.4	3.3	R	
1372	3.8	6.6	7.0	5.6	R	
1517	2.2	9.0	8.7	4.5	R	
1396	3.3	3.1	8.6	3.5	R	
1257	3.2	5.1	7.7	7.6	R	
1327	2.7	9.4	6.9	2.4	R	
1246	3.0	3.9	7.7	8.8	R	
1314	3.7	8.0	9.2	2.9	R	
1266	3.2	4.5	6.1	4.1	R	
1409	3.0	9.4	8.1	3.3	R	
1240	3.4	8.3	7.6	2.8	R	
1379	3.5	6.4	7.4	9.6	A	
1411	3.0	6.1	8.6	9.3	R	
1317	3.9	8.4	6.1	7.9	A	
1551	2.4	5.3	6.4	3.7	R	
1432	3.2	8.1	8.0	3.2	R	
1553	3.2	8.8	8.5	2.1	R	
1573	2.2	8.2	7.6	5.6	R	
1473	3.9	9.5	6.5	6.8	A	
1533	2.9	6.8	8.4	5.5	R	

Features

SAT	GPA	Essay	Rec.		Extra-C.	Decision
1309	3.0	7.3	8.4	6.6	R	
1531	2.4	4.3	8.2	7.4	R	
1335	2.9	4.9	6.7	2.6	R	
1313	2.2	4.1	6.1	6.1	R	
1441	3.4	6.2	8.1	8.7	R	
1219	2.7	4.8	8.4	9.7	R	
1480	3.7	6.1	7.3	5.7	R	
1208	2.3	7.5	6.8	7.7	R	
1328	3.1	4.0	8.4	7.7	R	
1286	3.8	8.7	7.5	9.2	A	
1332	4.0	7.2	9.1	6.0	A	
1374	3.7	9.1	8.8	5.4	R	
1534	2.8	5.0	7.1	7.8	R	
1498	2.6	9.1	7.6	6.1	R	
1433	3.4	3.5	8.9	4.5	R	
1515	3.9	4.3	8.4	4.6	R	
1233	3.9	5.1	8.7	7.8	A	
1521	2.2	6.9	8.7	8.6	R	
1591	3.0	4.6	7.0	8.3	R	
1578	2.0	3.7	9.3	5.2	R	
1446	2.8	3.7	8.7	3.1	R	

Feature Scaling

- The range of the features in a dataset can vary widely.
- The feature with a broad range of values tends to dominate the operations of the algorithm.
- Therefore, the range of all features should be scaled to a similar range so that each feature contributes approximately evenly to the results.

Normalization

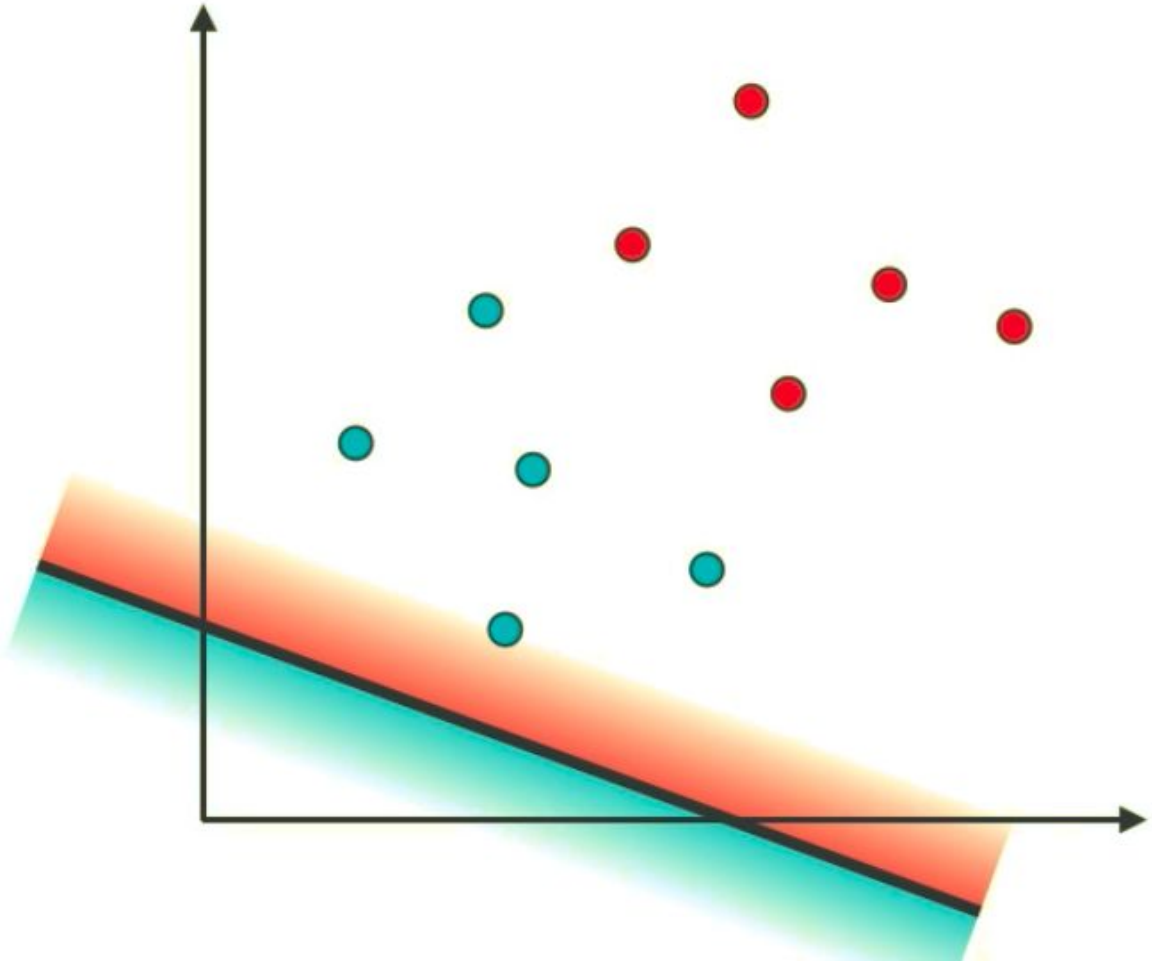
- Normalization is a scaling technique in which values are shifted and rescaled so that they end up ranging between 0 and 1. It is also known as Min-Max scaling. Here's the formula for normalization:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

- Since essay, rec. letters and extra-curricular activities data have been in the same range (0.0-10.0), we can scale SAT (1200-1600) and GPA (0.0-4.0) to the same range (0.0-10.0) instead of (0.0-1.0).

Perceptron Algorithm

Perceptron algorithm



Step 1: Start with a random line of equation $ax + by + c = 0$

Step 2: Pick a large number. **1000** (number of repetitions, or epochs)

Step 3: Pick a small number. **0.01** (learning rate)

Step 4: (repeat **1000** times)

- Pick random point **(p,q)**
- If point is correctly classified
  - Do nothing
- If point is blue, and $ap+bq+c > 0$
  - Subtract $0.01\mathbf{p}$ to a
  - Subtract $0.01\mathbf{q}$ to b
  - Subtract 0.01 to c
- If point is, red and $ap+bq+c < 0$
  - Add $0.01\mathbf{p}$ to a
  - Add $0.01\mathbf{q}$ to b
  - Add 0.01 to c

Step 5: Enjoy your line!

```
weights[2]*x + weights[1]*y + weights[0] = 0
weights[2] = randomFloat()*10.0;
weights[1] = randomFloat()*10.0;
weights[0] = randomFloat()*10.0;
```

```
#define MAX_ITERATION    1000;
```

```
#define LEARNING_RATE     0.01;
```

```
while (iteration < MAX_ITERATION) {…}
```

```
#define NO_PATTERNS    800;
r = randomly pick array element index
0…(NO_PATTERNS-1) or 0…799
if (label[r] < output) {
  weights[2] -= LEARNING_RATE* x[r];
  weights[1] -= LEARNING_RATE* y[r];
  weights[0] -= LEARNING_RATE;
}
else if (label[r] > output) {
  weights[2] += LEARNING_RATE* x[r];
  weights[1] += LEARNING_RATE* y[r];
  weights[0] += LEARNING_RATE;
}
```

```
weights[2]*x + weights[1]*y + weights[0] = 0
```

Step 1: Start with a random line of equation ax + by + c = 0

Step 2: Pick a large number. **1000** (number of repetitions, or epochs)

Step 3: Pick a small number. **0.01** (learning rate)

Step 4: (repeat **1000** times)

- Pick random point **(p,q)**
- If point is correctly classified
  - Do nothing
- If point is blue, and ap+bq+c > 0
  - Subtract 0.01**p** to a
  - Subtract 0.01**q** to b
  - Subtract 0.01 to c
- If point is, red and ap+bq+c < 0
  - Add 0.01**p** to a
  - Add 0.01**q** to b
  - Add 0.01 to c

Step 5: Enjoy your line!

College Application Calculator Project

Q&A