Classification and Discrimination

Nicholas Brust

Looking to allocate a sample X into a population group.

Methods

Classification Methods

Methods for classification Include the following.

- Minimizing the ECM
- Minimizing the TPM
- Fishers Model
- Logistic Regression
- Deep Learning Models.

Fisher Discriminant Rule

An Allocation Rule Based on Fisher's Discriminant Function⁵ Allocate \mathbf{x}_0 to π_1 if $\hat{y}_0 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{\text{pooled}}^{-1} \mathbf{x}_0$ $\geq \hat{m} = \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{\text{pooled}}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)$ or $\hat{y}_0 - \hat{m} \geq 0$ Allocate \mathbf{x}_0 to π_2 if $\hat{y}_0 < \hat{m}$ or $\hat{y}_0 - \hat{m} < 0$

has carried in home of this is created which the first first when the

This is a really great first example of a basic discrimination/ thresholding function to discriminate data.

Discrimination Example SAS

bill the basele blace islaming haryons constraints the main islaming i	•	Two Sample Discriminant Analysis
del backtyrg: mini i backtyrg: del backtyrg: mini backtyrg: del backt		The DISCRIM Procedure
in protein is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi import is all as its grouppi in diagonal is all as its grouppi i		Total Sample Size 46 DF Total 45
distance ()		Variables 4 DF Within Classes 44
0.53 0.51 1.07 0.55 0.52 0.64 0.52 0.65 0.55		Classes 2 DF Between Classes 1
-0.58 -0.31 1.51 0.56 0 -0.59 -0.59 1.56 0.57 0		
c.00		Number of Observations Read 46
•••••••••••••••••••••••••••••		
-0.07 1.56 0.57 0 -0.4 0.17 0.46 0.17 0.46 0.17 0.46 0.18 0 -0.4 0.17 0.54 0.17 0.54 0.17 0.46 0.18 0 -0.4 0.17 0.54 0.18 0 0 0.18 0.18 0 0 0 0.18 0.18 0 0 0 0.18 0.18 0 0 0 0.18 0.18 0.18 0		
-0.07 0.11 -0.07 0.28 0 0.04 0.04 1.37 0.34 0.48 0 0.05 0.02 1.37 0.34 0 <		
0.64 0.01 1.3 0.0.7 1.0 0.01 1.3 0.0.7 1.0 0.77 0.05 1.37 0.04 0 0.21 0.30 0.10 0 0.78 0.02 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.02 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.05 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.05 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.05 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.05 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.05 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.05 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.05 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.05 1.31 0.05 0 0.21 0.03 0.10 0 0.78 0.05 1.10 0.55 0.10 0 0.21 0.03 0.10 0 0.78 0.05 1.10 0.10 0 0.21 0.03 0.10 0 0.78 0.05 1.10 0.10 0 0.21 0.03 0.10 0 0.78 0.05 1.10 0.10 0 0.21 0.03 0.10 0 0.78 0.05 1.10 0.05 0.10 0 0.21 0.03 0.10 0 0.78 0.05 1.10 0.05 0.10 0 0.21 0.03 0.10 0 0.78 0.05 0.10 0 0.21 0.03 0.10 0 0.78 0.05 0.10 0 0.21 0.03 0.10 0 0.78 0.05 0.10 0 0.21 0.03 0.10 0 0.78 0.05 0.11 0 0.21 0.05 0.10 0 0.78 0.05 0.1 0 0.21 0.05 0.1 0 <t< th=""><th></th><th></th></t<>		
-0.07 -0.06 1.37 0.40 0 -0.03 -0.30 0.38 0 -0.03 -0.23 0.048 0.048 0.048 -0.04 -0.05 0.048 0.048 0.048 0.048 -0.04 -0.05 0.048		
corr corr<		0 0 21 21,0000 0.456522 0.500000
-0.14 1.42 0.43 0.43 0.43 0.43 0.44 0.43 0.44		1 1 25 25.0000 0.543478 0.500000
-0.32 -0.33 0.18 0 -0.32 -0.32 0.18 0.56 0 0.11 0.56 0.01 0.18 0.01 0.18 0.01 0.18 0.01 0.18 0.01 0.18 0.01 0.18 0.01 0.18 0.01 0.18 0.01 0.18 0.01 0.18 0.01 0.18 0.01 0.01 0.10 0.14 0.01 0.14 0.01 0.14 0.01 0.14 0.01 0.14 0.01 0.14 0.01 0.14 0.01 0.14 0.01 0.14 0.01 0.15 0.05 0.01 <		
0.07 0.13 0.25 0.31 0.05		
0.10 0.21 5 0.70 0 0.28 0.21 10 0.06 0 0.11 10 0.06 0 0.51 0.05 1 0 0.15 0.05 1.16 0.05 0 0 0.15 0.05 1.16 0.05 0 0 0.15 0.05 1.16 0.05 0 0 0.15 0.05 1.16 0.05 0 0 0.15 0.05 1.16 0.05 0 0 0.15 0.05 1.16 0.05 0 0 0.15 0.05 2.16 0.05 1 0 0.15 0.05 2.16 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0 0.15 0.05 2.15 0.05 1 0		Two Sample Discriminant Analysis
-0.23 -0.23 1.19 0.56 0 -0.13 0.05 0.03 0.05		The Discouter and the second sec
0.15 0.05 1.88 0.27 0 1.99 0.38 0		
0.11 1.99 0.38 0 0.08 0.63 1.64 0.55 0 0.12 0.11 1.42 0.55 0 0.13 0.15 0.45 0 0 0.55 0 0.12 0.11 1.44 0.17 0 0 0.55 0.05 1.65 0.05 0.		
-0.08 1.51 0.42 0 0.01 0.05 0.01 0.17 0.07 <		
0.10 1.28 0.25 0.10 1.28 0.25 0.10 0.12 0.11 1.11 0.17 0 0.00 0.00 1.00 0.00 <t< th=""><td>-0.08 -0.08 1.51 0.42 0</td><td></td></t<>	-0.08 -0.08 1.51 0.42 0	
0.10 1.25 0.10 1.25 0.10 0.24 0.10 0.24 0.10	0.05 0.03 1.68 0.95 0	
1.1 1.1	0.01 0.00 1.26 0.60 0	
0.10 0.10 1.12 0.13 0.10 0.11	0.12 0.11 1.14 0.17 0	
0.08 0.02 2.01 0.53 1 0.18 0.17 0.55 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 1 1 0.05 0.01 0.01 1		x4 0.0041473810 0.0034414286 0.0327816667 0.0445790476
0.18 0.11 3.27 0.35 1 0.19 0.05 2.25 0.33 1 0.22 0.07 4.24 0.63 1 0.22 0.07 4.24 0.63 1 0.20 0.02 0.25 0.33 1 0.22 0.07 4.24 0.63 1 0.20 0.02 0.25 0.33 1 0.22 0.07 1.80 0.52 1 0.15 0.05 2.31 0.25 1 0.15 0.05 2.31 0.25 1 0.16 0.05 0.16 1.88 0.08 1.89 0.08 1.89 0.08 1.89 0.08 1.89 0.08 1.89 0.08 1.89 0.08 0.08 1.89 0.08 0.08 0.08 0.08 0.08 0.08 0.08 0		
0.38 0.11 3.27 0.35 1 0.19 0.05 2.25 0.33 1 0.10 0.05 4.45 0.69 1 0.20 0.02 2.05 0.35 1 0.20 0.02 2.05 0.35 1 0.20 0.02 2.05 0.35 1 0.20 0.02 2.05 0.35 1 0.10 0.05 2.1 0 0.14 0.07 1.60 0.52 1 0.14 0.07 1.60 0.52 1 0.14 0.07 1.60 0.32 1 0.38 1 0.38 0.56 1.31 0.38 1 0.38 0.56 1.31 0.38 1 0.38 0.56 1.31 0.38 1 0.38 0.05 1.43 0.38 1 0.48 0.05 1.43 0.38 1 0.58 0.14 1.42 0.44 1 0.58 0.04 1.59 0.30 1 0.48 0.59 1.24 0.18 1 0.58 0.04 1.59 0.30		aroup = 1. DF = 24
0.32 0.07 4.24 0.63 1 0.33 0.68 1.45 0.69 1 0.31 0.05 2.55 0.69 1 0.02 0.06 2.35 0.69 1 0.02 0.06 2.35 0.40 1 0.12 0.07 2.61 0.52 1 0.14 0.07 2.61 0.52 1 0.14 0.07 2.61 0.52 1 0.15 0.62 2.33 0.26 1 0.14 0.07 2.61 0.52 1 0.14 0.07 2.61 0.52 1 0.15 0.65 1 0.14 0.07 2.61 0.52 1 0.15 0.65 0.13 1 0.29 0.68 0.13 1 0.29 0.68 0.13 1 0.29 0.68 0.13 1 0.29 0.68 0.13 1 0.20 0.08 5.06 0.13 1 0.20 0.20 0.20 0.20 0.20 0.20 0.20 0.20		
0.32 0.07 4.24 0.63 1 0.31 0.05 2.52 0.69 1 0.12 0.05 2.52 0.69 1 0.12 0.05 2.52 0.69 1 0.12 0.05 2.52 0.69 1 0.12 0.05 0.55 1 0 0.05 0.05 1 0.12 0.05 0.55 1 0 0.05 0.05 1 0.14 0.07 2.60 0.52 1 0 0.06 0.68 1 0.14 0.07 0.68 0.55 1 0.06 0.06 1.80 0.68 1 0.14 0.07 2.68 0.63 1 0.07 1.68 0.68 1 0.07 1.68 0.68 1 0.07 1.68 0.68 1 0.07 1.68 0.68 1 0.07 1.68 0.68 1 0.06 0.68 0.68 1 0.06 0.68 0.68 0.68 0.68 0.68 0.68		x1 0.047051000 0.005607167 0.074830500 -0.005703500
0.12 0.02 1.25 0.25 0.35 1 -0.02 0.02 2.25 0.35 1 -0.02 0.02 2.35 0.40 1 -0.12 0.00 2.55 0.58 1 -0.02 0.02 2.35 0.40 1 -0.12 0.00 1.250 0.58 1 -0.02 0.02 2.35 0.40 1 -0.12 0.00 1.250 0.58 1 -0.12 0.00 1.250 0.58 1 -0.12 0.00 1.250 0.58 1 -0.12 0.00 1.250 0.058 1 -0.12 0.00 1.250 0.018 1 -0.55 0.11 4.252 0.48 1 -0.55 0.11 4.252 0.48 1 -0.55 0.11 4.252 0.48 1 -0.55 0.11 4.252 0.45 1 -0.12 0.00 1.24 0.03 1 -0.47 0.14 2.22 0.45 1 -0.20 0.08 1.99 0.30 1 -0.48 0.09 1.24 0.18 1 -0.20 0.05 0.11 4.252 0.44 1 -0.20 0.05 0.11 4.252 0.45 1 -0.20 0.05 0.11 4.252 0.45 1 -0.20 0.05 0.18 1 -0.20 0.05 0.18 1 -0.20 0.05 0.01 1 -0.48 0.09 1.24 0.18 1 -0.20 0.05 0.01 1 -0.20 0.05 0.00 1 -0.20 0.00 1 -0.20 0.00 0 -0.20 0 -0.2		
0.12 0.02 2.25 0.65 1 0.12 0.02 2.05 1 0.12 0.02 0.05 1.15 0.15 0.05 1.15 0.05 1 0.14 0.07 2.61 1.52 1 0.14 0.07 2.61 0.52 1 0.14 0.07 2.61 0.52 1 0.14 0.07 2.61 0.52 1 0.14 0.08 1.80 0.81 1 0.15 0.62 1.80 0.81 1 0.14 0.07 2.61 0.52 1 1 0.14 0.07 2.61 0.52 1 1 1 0.20 0.84 1		
-0.22 0.02 2.05 0.13 1 0.12 0.02 2.35 0.46 1 0.12 0.02 0.05 1 <th></th> <th></th>		
0.17 0.07 1.80 0.52 1 0.15 0.05 1.77 0.55 1 0.14 0.07 2.61 0.52 1 0.14 0.07 2.61 0.52 1 0.14 0.07 2.61 0.52 1 0.14 0.07 2.61 0.52 1 0.15 0.05		
0.15 0.05 2.17 0.55 1 0.14 -0.03 2.50 0.58 1 0.14 -0.03 2.64 0.25 1 0.14 -0.03 2.64 0.25 1 0.15 0.05 2.31 0.20 1 0.15 0.05 2.31 0.20 1 0.15 0.05 2.31 0.20 1 0.54 0.12 3.3 0.08 1 0.54 0.12 3.3 0.08 1 0.54 0.12 1.23 0.48 1 0.55 0.11 4.22 0.48 1 0.55 0.01 4.2 0.18 1 0.55 0.01 4.2 0.18 1 0.55 0.01 4.2 0.18 1 0.55 0.01 4.2 0.50 0 0.55 0.01 1 0.55 0.01 1 0.55 0.00 0.01 1 0.55 0.00 0.00 0 0.55 0.00 0.00 0 0.55 0.00 0.00 0 0.55 0.00 0.00 0 0.55		
-0.10 -0.01 2.50 0.55 1 0.14 -0.03 0.46 1		Two Sample Discriminant Analysis
0.14 -0.03 0.46 0.26 1 0.14 -0.07 2.61 0.25 1 0.15 0.06 2.23 0.56 1 0.15 0.06 2.23 0.56 1 0.15 0.06 2.23 0.50 1 0.20 0.06 1.84 0.38 1 0.54 0.12 1.33 0.48 1 0.54 0.18 1 0.54 0.18 1 0.54 0.18 1 0.54 0.18 1 0.54 0.18 1 0.55 0.11 4.29 0.46 1 0.50 0.01 4.29 0.48 1 0.50 0.01 1 0.48 0.09 1.24 0.18 1 0.50 0.04 5.06 0.13 1 ; proc discipling discound disc		
0.14 0.07 2.61 0.52 1 0.15 0.06 2.23 0.56 1 0.16 0.05 2.31 0.20 1 0.29 0.64 1 0.33 0.08 1 0.33 0.08 1 0.33 0.44 1 0.50 0.01 4.29 0.44 1 0.50 0.00 1.99 0.30 1 0.47 0.14 2.92 0.45 1 0.17 0.04 2.45 0.11 1 0.50 0.04 5.06 0.13 1 ; prov uiscing additional grad backgrad backgr		The DISCRIM Procedure
0.15 0.06 2.23 0.56 1 0.16 0.05 2.33 0.20 1 0.54 0.12 1.33 0.48 1 0.54 0.11 4.29 0.46 1 0.54 0.11 4.29 0.46 1 0.55 0.11 4.29 0.46 1 0.50 0.01 4.2 0.18 1 0.50 0.01 4.29 0.46 1 0.50 0.01 5.06 0.13 1 ; proc usclass group; var x1 x2 x3 x4; run;		Pooled Within-Class Covariance Matrix, DF = 44
0.16 0.05 2.31 0.20 1 0.29 0.06 1.84 0.38 1 0.39 0.00 3.01 0.47 1 0.30 0.00 3.01 0.47 1 0.48 0.09 1.24 0.18 0.18 0.09 1 0.48 0.09 0.00 0.00 0.00 0.00 0.00 0.00		Variable x1 x2 x3 x4
0.29 0.06 1.84 0.38 1 0.54 0.12 2.33 0.48 1 0.54 0.12 2.33 0.48 1 0.54 0.13 1 0.55 0.14 1.23 0.44 1 0.55 0.14 1.23 0.44 1 0.52 0.18 1 0.55 0.14 1.20 0.45 1 0.17 0.04 2.45 0.14 1 0.56 0.14 1.00 0.50 0.13 1 ; proc uscur duburder bankster under bankster		x1 0.0457228398 0.0175841039 0.05654996970017712814
0.54 0.11 2.33 0.48 1 0.33 0.04 0.17 1 0.48 0.09 1.24 0.18 1 0.55 0.14 1 0.55 0.14 1 0.55 0.14 1 0.58 0.04 5.06 0.13 1 ; proc usscience usethod=normal pool=yes manova wcov pcov listerr crosslisterr; var x1 x2 x3 x4; run;		x2 0.0175841039 0.0108425714 0.0165090000 0.0016653766
-0.33 -0.09 3.01 0.47 1 0.48 0.09 1.24 0.18 1 0.56 0.11 4.29 0.44 1 0.50 0.18 1.99 0.30 1 0.47 0.14 2.92 0.45 1 0.17 0.04 5.06 0.13 1 ; proc uscient data - DamAi up c method=normal pool=yes manova wcov pcov listerr crosslisterr; class group; var x1 x2 x3 x4; run;		x3 0.0565499697 0.0165090000 0.6456509697 0.0327004848
0.48 0.09 1.24 0.18 1 0.56 0.11 4.29 0.44 1 0.20 0.08 1.99 0.30 1 0.47 0.14 2.92 0.45 1 0.17 0.04 2.45 0.14 1 0.58 0.04 5.06 0.13 1 ; proc discriming data-ballstupe method=normal pool=yes manova wcov pcov listerr crosslisterr; var x1 x2 x3 x4; run; • • • • • • • • • • • • •		x4 -0017712814 0.0016653766 0.0327004848 0.0346528398
0.50 0.11 4.29 0.44 1 0.20 0.08 1.99 0.30 1 0.42 0.08 1.99 0.30 1 0.42 0.45 1.90 0.30 1 0.42 0.45 0.01 1 1 proc uscient data-participe method=normal pool=yes manova wcov pcov listerr crosslisterr; var x1 x2 x3 x4; run;		
0.20 0.08 1.99 0.30 1 0.47 0.14 2.92 0.45 1 0.17 0.04 2.95 1.04 1 0.58 0.04 5.06 0.13 1 proc unscrime data-ballituge method=normal pool=yes manova wcov pcov listerr crosslisterr; var x1 x2 x3 x4; run;		
<pre>0.47 0.14 2.92 0.45 1 0.17 0.04 2.45 0.14 1 0.58 0.04 5.06 0.13 1 ; proc uscrim data-bankrupt method=normal pool=yes manova wcov pcov listerr crosslisterr; class group; var x1 x2 x3 x4; run;</pre>		
<pre>0.17 0.04 2.45 0.14 1 proc discrimentation pool=yes manova wcov pcov listerr crosslisterr; proc discrimentation pool=yes manova wcov pcov lister pool=yes manova wcov pcov listerr crosslisterr; proc discrimentation pool=yes manova wcov pcov lister pool=yes manova</pre>	0.47 0.14 2.92 0.45 1	Natural Log of the Covariance Determinant of the
<pre>proc userium gata-baintupt method=normal pool=yes manova wcov pcov listerr crosslisterr; var x1 x2 x3 x4; run;</pre>	0.17 0.04 2.45 0.14 1	Matrix Rank Covariance Matrix
proc userim data-pankiupi method=normal pool=yes manova wcov pcov listerr crosslisterr; class group; var x1 x2 x3 x4; run;	0.58 0.04 5.06 0.13 1	4 -12.63389
Class group; var x1 x2 x3 x4; run;		
Class group; var x1 x2 x3 x4; run;	proc urserim data-pankrupt method=normal pool=yes manova wcov pcov listerr crosslisterr;	
Var x1 x2 x3 x4; Generalized Squared Distance to group. run; 0 0 1		
Val XI X2 X3 X4; run;	class group;	The DISCRIM Procedure
run;	var x1 x2 x3 x4:	Generalized Squared Distance to group
run; 1 362722 0		
1 842722 0	f run;	
	· · · · · · · · · · · · · · · · · · ·	1 3.62722 0



Our goal with the TPM rule is to use the distance from the mean of each group in order to place our sample.

This data is uses GPA and GMAT to determine the admission status of individuals applying to graduate school.

TPM Example Continued

X0 is a sample student with minimizing at D3 we find the student right on the edge of admission.

With
$$\mathbf{x}'_0 = [3.21, 497]$$
, the sample squared distances are
 $D_1^2(\mathbf{x}_0) = (\mathbf{x}_0 - \bar{\mathbf{x}}_1)' \mathbf{S}_{\text{pooled}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_1)$
 $= [3.21 - 3.40, 497 - 561.23] \begin{bmatrix} 28.6096 & .0158 \\ .0158 & .0003 \end{bmatrix} \begin{bmatrix} 3.21 - 3.40 \\ 497 - 561.23 \end{bmatrix}$
 $= 2.58$
 $D_2^2(\mathbf{x}_0) = (\mathbf{x}_0 - \bar{\mathbf{x}}_2)' \mathbf{S}_{\text{pooled}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_2) = 17.10$
 $D_3^2(\mathbf{x}_0) = (\mathbf{x}_0 - \bar{\mathbf{x}}_3)' \mathbf{S}_{\text{pooled}}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_3) = 2.47$

Logistic regression

The Method

This is an approach where at least some of the variables are qualitative and in its simplest form works well with data labeled in a binary fashion.

An example of this would be Male or Female. Or in code 0 or 1.

The Model

Where normally we would just use a linear model in this case we will consider the ratio of the classes: odds = p or 1-p



Deep Learning Networks

Neurons

A neuron is Simply a function applied to the data followed by an activation function.

A group of these functions applied to the data is called a layer.



Linear Neuron

• Applies a linear transformation to the data.

$$y = x \mathbf{A}^{\mathsf{T}} + \boldsymbol{\beta}$$

Activation Functions

 Relu applies a rectified linear unit function which takes the max between input value and 0.

• Sigmoid Function applies:







Groups of neurons strong together to either apply a linear Function(Linear, BiLinear, etc.), Activation function(Sigmoid, ReLU), or Convolutional(We will talk about this later).



Loss functions and Optimizers

Loss Functions:

- MSE Loss:
- Cross Entropy Loss
- L1 Loss:

Optimizers:

- Adam
- Adagrad
- Stochastic Gradient Descent

The Basic Algorithm

Forward Pass: Sends the data forward through the layers in order to get a prediction.

Backward Propagation: through the forward pass we calculate loss and then propagate it backwards through the network.

Repeat.

Logistics Regression and Neural Net on the Iris Dataset

The Data

THe Iris data set is a set of 150 observations of 3 different species of Iris plant, with Four Features(Sepal length and width, Petal length and Width). Using an 80/20 data split to construct a training and testing data set.



Logistic Regression Using SAS Demo

Results

Using the Iris dataset with the code shown to the left we created a model using the training data that was able to place each testing data in its proper place.

				104	LUGISTIC	recessie		
	<pre>surveyselect data = Iris out=shuffled rate=1 outorder=random;</pre>				Model Infor	rmation		
proc s	drveyselect data - 1115 out-sharried rate-1 outorder-random,		Deta	in Set		WOR	IK.TRAIN	
			Res	spense Va	riable	type		
					aponse Lev			
- T	rain:		Nee				dative logit	8
			Opt	limization	Technique	Fishe	ir's scoring	t.
SP	et shuffled (obs = 120);			Number	of Observat	tions Read	120	
1				Number	of Observat	tiens Used	120	
-1- T				Orde	Response	Profile		
)ata ⊺	est;					Frequenc		
	t shuffled (firstobs = 121 obs=150);				1 1	4		
Se	c shuffied (firstobs = 121 obs=150);				2 2	3		
un;					3 3	3		
un,		Probab	-Eldes m			d over the k		~
						ence Status data points d		
nroc e	gplot Data = Iris;		0	orrpeite se	panason of a	sea points d	seacoid.	
	5							
scatte	<pre>er x = sw y = pw / group= type markerattrs=(symbol = circlefilled size = 10);</pre>		Score			ional Odds /		
				Chi-8q		4	Pr > CNS	
kaxis .	label="Sepal Width";			0.5	420	•	0.944	-
					Model Fit St	adiation .		
avis	label="Petal Width";		Criteric	an Interc		Intercept ar	nd Covaria	
unis	inder i con nadni j		AIC		267.518		13.0	į
un;			sc		273.093		30.	3
,			-2 Log I	L	263.518		10	k
			T	Feating City	shal Null Hy	pothesis: B	BETA+0	
mor 1	ogistic data=train outmodel=irislr;		Test		Chi-Sq.	uare DF	Py > CNS	-
				ihood Rati			×.000	
nodel	type = pw sw sl pl;		Score		112.0		<.000	
	cite - bu an ar bri		Wald		4.6	6113 4	0.329	N
run;			A	nalysis of I		Relihood E	atimates Wald	
		Parama	ber	OF Est	imate	Error Ch	u-Square	
		Intercep	8 1			4.9440	0.0707	
proc 1	ogistic inmodel=irislr;	Intercep	1 2	1 54		3.8217 9.4814	0.3961	
		10				9.0510	1,2190	
score	data=Test(drop=type) out=score;	2				4,9454	1,4677	
		pl		1 -12		7.8707	2.3355	
·un;					olda Ratio E	- Henrice		
						95% 1		
			ER		et Estimate	Confident	ce Limits	
data f	inal:		pre		+0.001		64.501 >000.000	
			-		416.481		>900.900	
me	rge score Test;		-		<0.001		29,904	
11	I type-type = 0 Then answer = 1;	Asso	ciation	of Predicts	ed Probabili	ities and Ob	aserved Re	
		Pere	ent Cons	cordant	10	0.0 Serne	ers' D	
	se answer = 0;	Pero	ent Disc	ordent		0.0 Game		
e1								
el		Pero	ent Tied			0.0 Tara-a		

Shallow Neural Network Using Python Demo

Results

Using a neural net I was able to create the same results using 4 linear layers and the relu activation function. Using those we successfully placed most of the data into the right class.

Computer Vision

Image Data: What an Image is made of.



Image data is high dimensional data with many features and is an interesting data type when it comes to classifications.

Feature Extraction(A New Type of Layer)

How do we teach a computer what is in an image. It all starts with methods of feature extraction. We have seen this before with PCA which can be done on images. However we have other Methods of extracting features.

Convolutional Neural Nets

Convolutional Layer: Breaks up our Image into smaller images and pulls out features.

Pooling Layer: Reduces Dimensions by combining outputs of CL.

Dense Layer: Is similar to our original NN which produces an answer to the question.



Difficulties of Computer Vision

Computer VIsion takes a lot of data in order to have good results.

It can also take a lot of time in order completely create a brand new model. That is why it is a very common practice to use another model to start and edit the last Layer. This is also known as transfer learning. Or applying a model pre created to complete a new task.

Any Questions?