

Elijah Andrushenko

CPTS 434

Introduction to Neural Networks & Applications

11-12-2019

Homework 9

1.

Chromosome Fitness after 100 Generations

0.97883

NaiveBayes All Attributes

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	432	97.2973 %
Incorrectly Classified Instances	12	2.7027 %
Kappa statistic	0.9362	
Mean absolute error	0.0258	
Root mean squared error	0.1539	
Relative absolute error	6.1594 %	
Root relative squared error	33.6744 %	
Total Number of Instances	444	

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	0.971	0.023	0.990	0.971	0.981	0.937	0.996	0.998
	2							
	0.977	0.029	0.935	0.977	0.956	0.937	0.996	0.989
	4							

Weighted Avg. 0.973 0.025 0.974 0.973 0.973 0.937 0.996 0.996

==== Confusion Matrix ====

a b <-- classified as

303 9 | a = 2

3 129 | b = 4

Top 5 Attributes

==== Attribute Selection on all input data ====

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 10 class):

Information Gain Ranking Filter

Ranked attributes:

0.674 2 uniformity.of.cell.size

0.634 3 uniformity.of.cell.shape

0.613 6 bare.nuclei

0.612 7 bland.chromatin

0.496 5 single.epithelial.cell.size

0.468 8 normal.nuclei

0.459 4 marginal.adhesion

0.429 1 clump.thickness

0.184 9 mitoses

Selected attributes: 2,3,6,7,5,8,4,1,9 : 9

Thus the top five are

1. uniformity.of.cell.size
2. uniformity.of.cell.shape
3. bare.nuclei
4. bland.chromatin
5. single.epithelial.cell.size

Naïve Bayes Top 5 Attributes

==== Run information ====

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: breast.cancer-weka.filters.unsupervised.attribute.Remove-R1,4,8-9

Instances: 444

Attributes: 6

uniformity.of.cell.size
uniformity.of.cell.shape
single.epithelial.cell.size
bare.nuclei
bland.chromatin
class

Test mode: 10-fold cross-validation

Time taken to build model: 0 seconds

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	433	97.5225 %
Incorrectly Classified Instances	11	2.4775 %

Kappa statistic	0.9416
Mean absolute error	0.029
Root mean squared error	0.1552
Relative absolute error	6.9419 %
Root relative squared error	33.9455 %
Total Number of Instances	444

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	0.971	0.015	0.993	0.971	0.982	0.942	0.994	0.998
	0.985	0.029	0.935	0.985	0.959	0.942	0.994	0.984
Weighted Avg.	0.975	0.019	0.976	0.975	0.975	0.942	0.994	0.993

=== Confusion Matrix ===

```

a  b  <-- classified as
303  9 |  a = 2
  2 130 |  b = 4
2.

```

Naïve Bayes All Attributes

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes
Relation: leukemia gene expression data names changed
Instances: 72
Attributes: 151
[list of attributes omitted]

Test mode: 10-fold cross-validation

Time taken to build model: 0.01 seconds

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	68	94.4444 %
Incorrectly Classified Instances	4	5.5556 %
Kappa statistic	0.88	
Mean absolute error	0.0556	
Root mean squared error	0.2357	
Relative absolute error	11.6559 %	
Root relative squared error	48.2804 %	
Total Number of Instances	72	

==== Detailed Accuracy By Class ====

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
	1.000	0.143	0.917	1.000	0.957	0.886	0.942	0.933
ALL								
	0.857	0.000	1.000	0.857	0.923	0.886	0.977	0.974
AML								
Weighted Avg.	0.944	0.087	0.949	0.944	0.944	0.886	0.956	0.949

==== Confusion Matrix ====

a b <-- classified as

44 0 | a = ALL

4 24 | b = AML

Top 5 Attribute Selection

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval

Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1

Relation: leukemia gene expression data names changed

Instances: 72

Attributes: 151

[list of attributes omitted]

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 151 leukemia_type):

Information Gain Ranking Filter

Ranked attributes:

0.609 143 MPO from Human myeloperoxidase gene, exons 1-4./ntype=DNA /annot=exon

0.607 124 TALDO Transaldolase

0.559 145 Low-Mr GTP-binding protein (RAB31) mRNA

0.555 100 LRPAP1 Low density lipoprotein-related protein-associated protein 1 (alpha-2-macroglobulin receptor-associated protein 1

0.555 148 LPAP gene

0.555 147 OBF-1 mRNA for octamer binding factor 1

0.55 150 ALCAM Activated leucocyte cell adhesion molecule

0.517 121 GB DEF = Myosin-IE

0.505 125 Transmembrane protein

0.505 146 KIAA0184 gene, partial cds

0.492 135 FES Feline sarcoma (Snyder-Theilen) viral (v-fes)/Fujinami avian sarcoma (PRCII) viral (v-fps) oncogene homolog

0.478 142 RNH Ribonuclease/angiogenin inhibitor

0.477 106 Clone 23721 mRNA sequence

0.477 144 Very-long-chain acyl-CoA dehydrogenase (VLCAD)

0.465 68 SMT3B protein

0.451 62 Keratin 10 type I intermediate filament (KRT10) mRNA

0.446 134 Heterochromatin protein HP1Hs-gamma mRNA

0.444 140 MHC class I-related protein mRNA

0.44 74 Thymopoietin beta mRNA

0.435 99 AKT1 V-akt murine thymoma viral oncogene homolog 1

0.423 110 LMNA Lamin A

0.423 129 P4HB Procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), beta polypeptide (protein disulfide isomerase; thyroid hormone binding protein p55)

0.416 123 Skeletal muscle LIM-protein SLIM1 mRNA

0.414 80 PCBD 6-pyruvoyl-tetrahydropterin synthase/dimerization cofactor of hepatocyte nuclear factor 1 alpha (TCF1)

0.407 42 TCF3 Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)2

0.399 116 Tetracycline transporter-like protein mRNA

0.399 149 HOXB3 Homeo box B3

0.389 136 IGB Immunoglobulin-associated beta (B29)

0.386 87 CHRNA7 Cholinergic receptor, nicotinic, alpha polypeptide 7

0.386 137 KIAA0212 gene

0.384 93 CAB3b mRNA for calcium channel beta3 subunit

0.381 130 GLRX Glutaredoxin (thioltransferase)

0.371 95 KIAA0128 gene, partial cds

0.366 131 Metargidin precursor mRNA

0.36 127 OS-9 precurosor mRNA

0.359 34 ALDR1 Aldehyde reductase 1 (low Km aldose reductase)

0.356 132 BB1

0.351 138 Uridine phosphorylase

0.349 90 PHOSPHATIDYLINOSITOL

0.347 113 GB DEF = Homeodomain protein HoxA9 mRNA

0.347 101 SNRPN Small nuclear ribonucleoprotein polypeptide N

0.343 67 TOP2B Topoisomerase (DNA) II beta (180kD)

0.336 84 PLECKSTRIN

0.329 76 LEUKOCYTE ELASTASE INHIBITOR

0.328 92 GB DEF = GTP-binding protein (RAB3B) mRNA

0.322 122 Activin type II receptor

0.322 72 GTF2E2 General transcription factor TFIIE beta subunit, 34 kD

0.322 118 PI Protease inhibitor 1 (anti-elastase), alpha-1-antitrypsin

0.322 60 ALDH7 Aldehyde dehydrogenase 7

0.316 54 PTH2 parathyroid hormone receptor mRNA

0.316 139 THYMOSIN BETA-10

0.316 119 Epb72 gene exon 1

0.314 102 Amyloid precursor protein-binding protein 1 mRNA

0.313 69 Calcineurin A catalytic subunit [human, testis, mRNA, 2134 nt]

0.31 141 TFIID subunit TAFII55 (TAFII55) mRNA

0.31 8 GLUTATHIONE S-TRANSFERASE, MICROSOMAL

0.31 111 LMP2 gene extracted from H.sapiens genes TAP1, TAP2, LMP2, LMP7 and DOB

0.308 104 ADPRT ADP-ribosyltransferase (NAD⁺; poly (ADP-ribose) polymerase)

0.305 94 CD36 CD36 antigen (collagen type I receptor, thrombospondin receptor)

0.305 105 GB DEF = P85 beta subunit of phosphatidyl-inositol-3-kinase

0.305 50 Clone 22 mRNA, alternative splice variant alpha-1

0.295 10 ATP6C Vacuolar H⁺ ATPase proton channel subunit

0.295 96 SELL Leukocyte adhesion protein beta subunit2

0.295 83 Terminal transferase mRNA

0.294 133 HPrp18 mRNA

0.294 32 Nucleoside-diphosphate kinase

0.292 15 SPI1 Spleen focus forming virus (SFFV) proviral integration oncogene spi1

0.292 47 Transcriptional activator hSNF2b1

0.292 49 Oncoprotein 18 (Op18) gene

0.289 108 GB DEF = DNA for cellular retinol binding protein (CRBP) exons 3 and 4

0.289 65 CDC10 Cell division cycle 10 (homologous to CDC10 of *S. cerevisiae*)

0.289 37 LAMP2 Lysosome-associated membrane protein 2 {alternative products}

0.288 13 PPGB Protective protein for beta-galactosidase (galactosialidosis)

0.288 75 TCF12 Transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)

0.288 97 Nuclear Factor Nf-IL6

0.286 57 KIAA0022 gene

0.286 19 FAH Fumarylacetoacetate

0.284 114 CSF3R Colony stimulating factor 3 receptor (granulocyte)

0.284 120 Interleukin 8 (IL8) gene

0.283 126 HMG1 High-mobility group (nonhistone chromosomal) protein 1

0.283 64 INTERFERON GAMMA UP-REGULATED I-5111 PROTEIN PRECURSOR

0.283 79 PERIPHERAL-TYPE BENZODIAZEPINE RECEPTOR

0.281 128 Small GTP-binding protein mRNA

0.281 77 GRN Granulin

0.281 3 CD33 CD33 antigen (differentiation antigen)

0.277 88 GB DEF = Retinoblastoma susceptibility protein (RB1) gene, with a 3 bp deletion in exon 22 (L11910 bases 161855-162161)

0.274 4 DF D component of complement (adipsin)

0.274 63 GB DEF = Topoisomerase type II (Topo II) mRNA, partial cds

0.272 17 Lysozyme gene (EC 3.2.1.17)

0.27 5 RNS2 Ribonuclease 2 (eosinophil-derived neurotoxin; EDN)

0.27 109 ACADM Acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain

0.265 22 MYL1 Myosin light chain (alkali)

0.265 89 IL7R Interleukin 7 receptor

0.265 112 SERINE/THREONINE PROTEIN PHOSPHATASE 2B CATALYTIC SUBUNIT, BETA ISOFORM

0.265 107 Thymopoietin (TMPO) gene

0.262 30 Azurocidin gene

0.262 29 FTL Ferritin, light polypeptide

0.262 70 PRKCD Protein kinase C, delta

0.261 117 Putative enterocyte differentiation promoting factor mRNA, partial cds

0.259 28 MB-1 gene

0.254 40 SELL Leukocyte adhesion protein beta subunit1

0.254 48 INTERLEUKIN-8 PRECURSOR

0.254 21 Leukotriene C4 synthase (LTC4S) gene

0.254 16 LYZ Lysozyme1

0.254 98 Phosphotyrosine independent ligand p62 for the Lck SH2 domain mRNA

0.254 103 DCK Deoxycytidine kinase

0.254 31 TIMP2 Tissue inhibitor of metalloproteinase 2

0.254 81 Heat Shock Protein, 70 Kda (Gb:Y00371)

0.252 51 CYSTATIN A

0.252 18 LYZ Lysozyme2

0.245 25 Liver mRNA for interferon-gamma inducing factor(IGIF)

0.241 85 RHD Rhesus blood group, D antigen

0.241 53 Hunc18b2

0.24 14 HKR-T1

0.239 115 Hepatocyte growth factor-like protein gene

0.239 27 MPO Myeloperoxidase

0.239 71 Inducible protein mRNA

0.238 44 ITGAX Integrin, alpha X (antigen CD11C (p150), alpha polypeptide)

0.234 39 Quiescin (Q6) mRNA, partial cds

0.232 45 PFC Properdin P factor, complement

0.23 12 LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog

0.228 1 Zyxin

0.228 56 MANB Mannosidase alpha-B (lysosomal)

0.228 58 SPTAN1 Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)

0.228 41 CCND3 Cyclin D3

0.223 59 VIL2 Villin 2 (ezrin)

0.218 35 ME491 gene extracted from H.sapiens gene for Me491/CD63 antigen

0.217 86 Transcriptional activator hSNF2b2

0.216 78 CRYZ Crystallin zeta (quinone reductase)

0.216 73 NADH:ubiquinone oxidoreductase subunit B13 (B13) mRNA

0.216 91 PGD Phosphogluconate dehydrogenase

0.216 46 PROBABLE G PROTEIN-COUPLED RECEPTOR LCR1 HOMOLOG

0.21 24 ARHG Ras homolog gene family, member G (rho G)

0.208 61 ZNF91 Zinc finger protein 91 (HPF7, HTF10)

0.207 66 26-kDa cell surface protein TAPA-1 mRNA

0.205 33 Macmarcks

0.204 36 GPX1 Glutathione peroxidase 1

0.203 55 MSH2 DNA repair protein MSH2

0.203 26 TCF3 Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)1

0.202 82 ERC-55 mRNA

0.201 43 KAI1 Kangai 1 (suppression of tumorigenicity 6, prostate; CD82 antigen (R2 leukocyte antigen, antigen detected by monoclonal and antibody IA4))

0.191 7 APLP2 Amyloid beta (A4) precursor-like protein 2

0.191 38 C-myb gene extracted from Human (c-myb) gene, complete primary cds, and five complete alternatively spliced cds

0.19 20 RABAPTIN-5 protein

0.188 6 CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)

0.185 23 ALDOA Aldolase A

0.184 52 14-3-3 PROTEIN TAU

0.18 2 PRG1 Proteoglycan 1, secretory granule

0.173 11 PROTEASOME IOTA CHAIN

0.168 9 CTSD Cathepsin D (lysosomal aspartyl protease)

Selected attributes:

143,124,145,100,148,147,150,121,125,146,135,142,106,144,68,62,134,140,74,99,110,129,123,80,42,116,149,136,87,137,93,130,95,131,127,34,132,138,90,113,101,67,84,76,92,122,72,118,60,54,139,119,102,69,141,8,111,104,94,105,50,10,96,83,133,32,15,47,49,108,65,37,13,75,97,57,19,114,120,126,64,79,128,77,3,88,4,63,17,5,109,22,89,112,107,30,29,70,117,28,40,48,21,16,98,103,31,81,51,18,25,85,53,14,115,27,71,44,39,45,12,1,56,58,41,59,35,86,78,73,91,46,24,61,66,33,36,55,26,82,43,7,38,20,6,23,52,2,11,9 : 150

Thus the top 5 are

1. MPO from Human myeloperoxidase gene, exons 1-4./ntype=DNA /annot=exon
2. TALDO Transaldolase
3. Low-Mr GTP-binding protein (RAB31) mRNA
4. LRPAP1 Low density lipoprotein-related protein-associated protein 1 (alpha-2-macroglobulin receptor-associated protein 1
5. LPAP gene

Naïve Bayes Top 5 Attributes

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 65 90.2778 %

Incorrectly Classified Instances	7	9.7222 %
Kappa statistic	0.7914	
Mean absolute error	0.0993	
Root mean squared error	0.3079	
Relative absolute error	20.8388 %	
Root relative squared error	63.0753 %	
Total Number of Instances	72	

=== Detailed Accuracy By Class ===

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
ALL	0.955	0.179	0.894	0.955	0.923	0.795	0.952	0.955
AML	0.821	0.045	0.920	0.821	0.868	0.795	0.943	0.920
Weighted Avg.	0.903	0.127	0.904	0.903	0.902	0.795	0.948	0.941

==== Confusion Matrix ====

a b <-- classified as

42 2 | a = ALL

5 23 | b = AML

Results

For Breast Cancer Data Accuracy for all attributes was 97.3 % and for top 5 attributes it was 97.5% so we saw a gain in accuracy using on the top 5.

But in the Leukemia Data Set for all attributes the accuracy 94.4 % and for top 5 attributes it was 90.3% so we saw a loss in accuracy using the top 5.