1.1

当A0的属性值在90-110时，class0的可能性比较大。当A5接近于0时，class0的可能性较大。有些属性呈现正太分布，中心点在。。。。为什么？？？

1.2

a 画表格

b 是的，它反映两变量间线性相关关系，绝对值越大，相关程度也就越大

c 不一定，只能反应线性关系。有可能相关程度较大，但不是线性相关，系数的绝对值就会比较小

coef\_list: [0.4912, 0.0874, 0.2273, 0.2074, 0.1077, 0.1857, 0.0763, 0.3045, 0.2403]

1.3

数学公式推理

1.4

1.5

var\_list: [[5, 14769.0983], [0, 1006.4338], [6, 384.4417], [4, 254.105], [8, 139.6277], [7, 63.5047], [1, 16.7234], [2, 11.596], [3, 0.1059]]

sumofvar: 16645.6365

cum\_var\_ratio: [0.0605, 0.0615, 0.0622, 0.0622, 0.0775, 0.9648, 0.9879, 0.9917, 1.0000]

corr\_coefs: [[0.38559301420690023, -0.04582516082225113, -0.05705150513641671, 0.18579338880949564, 0.45924021835727835, 0.9996780135176875, 0.10057182197062049, 0.23230210476652127, -0.0015740705358918323], [-0.9142889454225337, -0.09077312003733048, -0.22546263478229117, -0.07987177012105025, 0.0971645653932254, 0.024084687085466953, -0.25539200876602997, -0.17260644028984368, -0.373443768831269]]

1.6

var\_list: [2.17949748 1.70751811 1.05032771 1.00190162 0.87165367 0.7420901

0.65742709 0.40969079 0.39276897]

sumofvar: 9.012875536480692

cum\_var\_ratio: [0.24182043 0.43127363 0.54780999 0.65897336 0.75568541 0.83802208

0.91096518 0.95642135 1. ]

corr\_coefs: [[0.6006713897085362, 0.05729590985040144, 0.26795203223120007, 0.3656554651376427, 0.6230382061060178, 0.6298568548408799, 0.5228507215944558, 0.6511607443951044, 0.3528644323802544], [0.17743225164359924, 0.1000368642956893, 0.7599827142218681, -0.2076433715300961, -0.4659904204832914, -0.36979586693725985, 0.22422808654293153, -0.16844950994503782, 0.7812491257116679]]

1.7

文字说明比较

1.8

the highest mean cross-validation accuracy: 0.7742857142857142

the value of C: 0.46415888336127775

the number of instances correctly classified: 75.0

classiﬁcation accuracy: 0.75

1.9

a:

mean vector: [27.020943396226414, 31.093207547169815]

covariance matrix: [[95.14113475 41.46999034]

[41.46999034 46.69341618]]

b:

1.11

mean of cross-validation accuracy: 0.7714285714285716

standard deviation of cross-validation accuracy: 0.04356557337707687

mean\_list: [0.6914285714285714, 0.7657142857142858, 0.7828571428571429, 0.7557142857142857, 0.7742857142857144, 0.77, 0.7628571428571429, 0.7557142857142857, 0.7685714285714286]

std\_list: [0.02976095236571379, 0.04637909451642268, 0.04664236490661289, 0.039486499888815016, 0.03953815000885159, 0.03736199094463434, 0.030772370326482883, 0.043892261416392034, 0.03901857223927904]

2.1

Train: min: 0.0 max: 1.0 mean: 0.17737719254368406 std: 0.3349824841708448

Test: min: 0.0 max: 1.0 mean: 0.17563382275987316 std: 0.3334625647782058

2.2

a

same

欧氏距离定义： 欧氏距离（ Euclidean distance）是一个通常采用的距离定义，它是在m维空间中两个点之间的真实距离

b

不应该改变，我们是对Xtrn进行训练，减的是Xtrn的均值，所以对Xtst进行测试时，也应该减去Xtrn的均值

2.3

2.4

2.5

Classification accuracy on training set: 0.916154

Classification accuracy on testing set: 0.722308

the numbers of misclassifications: [['L', 53], ['R', 48], ['I', 42], ['K', 38], ['N', 36], ['X', 35], ['E', 34], ['J', 33], ['A', 32], ['B', 30], ['G', 29], ['Z', 29], ['Q', 28], ['H', 27], ['U', 27], ['S', 26], ['V', 26], ['Y', 25], ['F', 23], ['W', 22], ['O', 18], ['D', 16], ['P', 13], ['M', 12], ['C', 10], ['T', 10]]

2.6

2.7

min\_cov: -0.0340630204459252

max\_cov: 0.1263636921707994

mean\_cov: 0.017079531291251494

min\_diag: 0.053432219716010185

max\_diag: 0.1263636921707994

mean\_diag: 0.08753841090172278

2.8

log\_likelihood: -1712612.7436858357

number of correctly classified instances in training: 7800

classification accuracy in training: 1.0

number of correctly classified instances in testing: 1803

classification accuracy in testing: 0.6934615384615385

2.9

2.10

acc\_training\_list: [1.0, 1.0, 1.0, 1.0]

acc\_testing\_list: [0.6934615384615385, 0.735, 0.7619230769230769, 0.8107692307692308]