Ziran Cao

Split table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Iteration | numerofPositiveSamplesInTraining | numberOfNegativeSamplesInTraininig | NumberofPositiveSamplesInDevelopment | NumberOfNegativeSamplesInDevelopment |
| 1 | 2230 | 1450 | 558 | 363 |
| 2 | 2231 | 1450 | 557 | 363 |
| 3 | 2231 | 1450 | 557 | 363 |
| 4 | 2230 | 1451 | 558 | 362 |
| 5 | 2230 | 1451 | 558 | 362 |

Iteration and probability

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Iteration | Pr(F1<=mui | spam) | Pr(F1>mui | spam) | Pr(F1<=mui | non-spam) | Pr(F1>mui | non-spam) | Pr(F2<=mui | spam) | Pr(F2>mui | spam) |
| 1 | 0.709 | 0.291 | 0.883 | 0.117 | 0.748 | 0.252 |
| 2 | 0.705 | 0.295 | 0.886 | 0.114 | 0.748 | 0.252 |
| 3 | 0.703 | 0.297 | 0.894 | 0.106 | 0.752 | 0.248 |
| 4 | 0.717 | 0.283 | 0.886 | 0.114 | 0.748 | 0.252 |
| 5 | 0.715 | 0.285 | 0.889 | 0.111 | 0.757 | 0.243 |

All the data are printed below, the table is too large to construct

Iteration:

1 [0.709, 0.748, 0.477, 0.979, 0.511, 0.655, 0.644, 0.697, 0.726, 0.634, 0.692, 0.564, 0.714, 0.874, 0.85, 0.55, 0.656, 0.672, 0.363, 0.799, 0.342, 0.952, 0.683, 0.671, 0.991, 0.992, 0.999, 0.985, 0.997, 0.992, 0.998, 0.999, 0.971, 0.995, 0.983, 0.942, 0.954, 0.983, 0.968, 0.894, 0.999, 0.994, 0.953, 0.984, 0.85, 0.986, 0.989, 0.99, 0.913, 0.74, 0.932, 0.452, 0.493, 0.803, 0.707, 0.586, 0.597]

1 [0.291, 0.252, 0.523, 0.021, 0.489, 0.345, 0.356, 0.303, 0.274, 0.366, 0.308, 0.436, 0.286, 0.126, 0.15, 0.45, 0.344, 0.328, 0.637, 0.201, 0.658, 0.048, 0.317, 0.329, 0.009, 0.008, 0.001, 0.015, 0.003, 0.008, 0.002, 0.001, 0.029, 0.005, 0.017, 0.058, 0.046, 0.017, 0.032, 0.106, 0.001, 0.006, 0.047, 0.016, 0.15, 0.014, 0.011, 0.01, 0.087, 0.26, 0.068, 0.548, 0.507, 0.197, 0.293, 0.414, 0.403]

1 [0.883, 0.932, 0.804, 0.997, 0.85, 0.904, 0.985, 0.94, 0.93, 0.86, 0.952, 0.695, 0.899, 0.955, 0.985, 0.941, 0.924, 0.898, 0.711, 0.987, 0.818, 0.994, 0.984, 0.986, 0.684, 0.739, 0.798, 0.848, 0.872, 0.84, 0.893, 0.925, 0.881, 0.924, 0.842, 0.836, 0.757, 0.98, 0.884, 0.914, 0.951, 0.896, 0.9, 0.907, 0.75, 0.849, 0.982, 0.932, 0.857, 0.618, 0.866, 0.902, 0.962, 0.938, 0.96, 0.927, 0.858]

1 [0.117, 0.068, 0.196, 0.003, 0.15, 0.096, 0.015, 0.06, 0.07, 0.14, 0.048, 0.305, 0.101, 0.045, 0.015, 0.059, 0.076, 0.102, 0.289, 0.013, 0.182, 0.006, 0.016, 0.014, 0.316, 0.261, 0.202, 0.152, 0.128, 0.16, 0.107, 0.075, 0.119, 0.076, 0.158, 0.164, 0.243, 0.02, 0.116, 0.086, 0.049, 0.104, 0.1, 0.093, 0.25, 0.151, 0.018, 0.068, 0.143, 0.382, 0.134, 0.098, 0.038, 0.062, 0.04, 0.073, 0.142]

2 [0.705, 0.748, 0.479, 0.979, 0.514, 0.652, 0.65, 0.703, 0.723, 0.63, 0.686, 0.568, 0.711, 0.879, 0.852, 0.539, 0.649, 0.668, 0.359, 0.795, 0.334, 0.95, 0.69, 0.665, 0.991, 0.99, 0.999, 0.983, 0.997, 0.991, 0.998, 0.999, 0.972, 0.994, 0.987, 0.949, 0.954, 0.983, 0.961, 0.892, 0.999, 0.995, 0.953, 0.983, 0.854, 0.987, 0.99, 0.991, 0.914, 0.736, 0.937, 0.436, 0.49, 0.796, 0.699, 0.576, 0.596]

2 [0.295, 0.252, 0.521, 0.021, 0.486, 0.348, 0.35, 0.297, 0.277, 0.37, 0.314, 0.432, 0.289, 0.121, 0.148, 0.461, 0.351, 0.332, 0.641, 0.205, 0.666, 0.05, 0.31, 0.335, 0.009, 0.01, 0.001, 0.017, 0.003, 0.009, 0.002, 0.001, 0.028, 0.006, 0.013, 0.051, 0.046, 0.017, 0.039, 0.108, 0.001, 0.005, 0.047, 0.017, 0.146, 0.013, 0.01, 0.009, 0.086, 0.264, 0.063, 0.564, 0.51, 0.204, 0.301, 0.424, 0.404]

2 [0.886, 0.928, 0.808, 0.996, 0.849, 0.915, 0.984, 0.942, 0.938, 0.856, 0.953, 0.701, 0.902, 0.959, 0.985, 0.942, 0.926, 0.898, 0.704, 0.988, 0.815, 0.992, 0.982, 0.983, 0.692, 0.739, 0.802, 0.853, 0.876, 0.842, 0.896, 0.926, 0.883, 0.926, 0.845, 0.836, 0.755, 0.983, 0.888, 0.915, 0.948, 0.9, 0.904, 0.906, 0.745, 0.846, 0.983, 0.937, 0.847, 0.622, 0.866, 0.905, 0.964, 0.941, 0.959, 0.936, 0.865]

2 [0.114, 0.072, 0.192, 0.004, 0.151, 0.085, 0.016, 0.058, 0.062, 0.144, 0.047, 0.299, 0.098, 0.041, 0.015, 0.058, 0.074, 0.102, 0.296, 0.012, 0.185, 0.008, 0.018, 0.017, 0.308, 0.261, 0.198, 0.147, 0.124, 0.158, 0.104, 0.074, 0.117, 0.074, 0.155, 0.164, 0.245, 0.017, 0.112, 0.085, 0.052, 0.1, 0.096, 0.094, 0.255, 0.154, 0.017, 0.063, 0.153, 0.378, 0.134, 0.095, 0.036, 0.059, 0.041, 0.064, 0.135]

3 [0.703, 0.752, 0.494, 0.981, 0.52, 0.65, 0.634, 0.701, 0.733, 0.627, 0.687, 0.57, 0.719, 0.888, 0.864, 0.557, 0.657, 0.679, 0.37, 0.808, 0.346, 0.95, 0.692, 0.665, 0.988, 0.989, 0.999, 0.984, 0.999, 0.991, 0.998, 0.999, 0.971, 0.994, 0.988, 0.947, 0.951, 0.984, 0.968, 0.901, 0.999, 0.994, 0.959, 0.986, 0.852, 0.987, 0.99, 0.992, 0.91, 0.735, 0.937, 0.458, 0.484, 0.801, 0.71, 0.585, 0.609]

3 [0.297, 0.248, 0.506, 0.019, 0.48, 0.35, 0.366, 0.299, 0.267, 0.373, 0.313, 0.43, 0.281, 0.112, 0.136, 0.443, 0.343, 0.321, 0.63, 0.192, 0.654, 0.05, 0.308, 0.335, 0.012, 0.011, 0.001, 0.016, 0.001, 0.009, 0.002, 0.001, 0.029, 0.006, 0.012, 0.053, 0.049, 0.016, 0.032, 0.099, 0.001, 0.006, 0.041, 0.014, 0.148, 0.013, 0.01, 0.008, 0.09, 0.265, 0.063, 0.542, 0.516, 0.199, 0.29, 0.415, 0.391]

3 [0.894, 0.932, 0.803, 0.998, 0.852, 0.907, 0.985, 0.944, 0.936, 0.863, 0.953, 0.702, 0.898, 0.959, 0.988, 0.939, 0.926, 0.903, 0.704, 0.987, 0.815, 0.992, 0.984, 0.984, 0.685, 0.739, 0.803, 0.855, 0.873, 0.848, 0.896, 0.926, 0.885, 0.925, 0.848, 0.835, 0.76, 0.985, 0.887, 0.916, 0.95, 0.896, 0.903, 0.905, 0.748, 0.853, 0.985, 0.935, 0.856, 0.623, 0.863, 0.908, 0.962, 0.931, 0.958, 0.93, 0.859]

3 [0.106, 0.068, 0.197, 0.002, 0.148, 0.093, 0.015, 0.056, 0.064, 0.137, 0.047, 0.298, 0.102, 0.041, 0.012, 0.061, 0.074, 0.097, 0.296, 0.013, 0.185, 0.008, 0.016, 0.016, 0.315, 0.261, 0.197, 0.145, 0.127, 0.152, 0.104, 0.074, 0.115, 0.075, 0.152, 0.165, 0.24, 0.015, 0.113, 0.084, 0.05, 0.104, 0.097, 0.095, 0.252, 0.147, 0.015, 0.065, 0.144, 0.377, 0.137, 0.092, 0.038, 0.069, 0.042, 0.07, 0.141]

4 [0.717, 0.748, 0.491, 0.981, 0.524, 0.674, 0.66, 0.7, 0.73, 0.636, 0.708, 0.572, 0.729, 0.878, 0.861, 0.56, 0.669, 0.686, 0.367, 0.807, 0.34, 0.943, 0.688, 0.669, 0.991, 0.992, 1.0, 0.984, 0.997, 0.99, 0.999, 0.999, 0.975, 0.995, 0.985, 0.948, 0.953, 0.98, 0.968, 0.906, 1.0, 0.993, 0.957, 0.981, 0.851, 0.986, 0.99, 0.993, 0.918, 0.744, 0.928, 0.455, 0.494, 0.802, 0.713, 0.596, 0.619]

4 [0.283, 0.252, 0.509, 0.019, 0.476, 0.326, 0.34, 0.3, 0.27, 0.364, 0.292, 0.428, 0.271, 0.122, 0.139, 0.44, 0.331, 0.314, 0.633, 0.193, 0.66, 0.057, 0.312, 0.331, 0.009, 0.008, 0.0, 0.016, 0.003, 0.01, 0.001, 0.001, 0.025, 0.005, 0.015, 0.052, 0.047, 0.02, 0.032, 0.094, 0.0, 0.007, 0.043, 0.019, 0.149, 0.014, 0.01, 0.007, 0.082, 0.256, 0.072, 0.545, 0.506, 0.198, 0.287, 0.404, 0.381]

4 [0.886, 0.93, 0.811, 0.997, 0.858, 0.907, 0.987, 0.942, 0.93, 0.865, 0.954, 0.704, 0.896, 0.961, 0.984, 0.938, 0.926, 0.9, 0.7, 0.989, 0.813, 0.992, 0.981, 0.983, 0.687, 0.735, 0.804, 0.855, 0.876, 0.843, 0.895, 0.927, 0.886, 0.926, 0.846, 0.836, 0.758, 0.983, 0.884, 0.913, 0.942, 0.897, 0.897, 0.909, 0.748, 0.844, 0.984, 0.936, 0.852, 0.62, 0.862, 0.9, 0.964, 0.937, 0.961, 0.935, 0.858]

4 [0.114, 0.07, 0.189, 0.003, 0.142, 0.093, 0.013, 0.058, 0.07, 0.135, 0.046, 0.296, 0.104, 0.039, 0.016, 0.062, 0.074, 0.1, 0.3, 0.011, 0.187, 0.008, 0.019, 0.017, 0.313, 0.265, 0.196, 0.145, 0.124, 0.157, 0.105, 0.073, 0.114, 0.074, 0.154, 0.164, 0.242, 0.017, 0.116, 0.087, 0.058, 0.103, 0.103, 0.091, 0.252, 0.156, 0.016, 0.064, 0.148, 0.38, 0.138, 0.1, 0.036, 0.063, 0.039, 0.065, 0.142]

5 [0.715, 0.757, 0.471, 0.978, 0.509, 0.662, 0.635, 0.705, 0.731, 0.647, 0.7, 0.575, 0.731, 0.88, 0.854, 0.546, 0.674, 0.676, 0.38, 0.795, 0.35, 0.949, 0.688, 0.673, 0.99, 0.99, 0.999, 0.983, 0.997, 0.992, 0.999, 0.999, 0.972, 0.994, 0.985, 0.946, 0.952, 0.981, 0.971, 0.893, 0.999, 0.994, 0.955, 0.98, 0.853, 0.985, 0.99, 0.99, 0.917, 0.735, 0.936, 0.447, 0.507, 0.81, 0.706, 0.584, 0.604]

5 [0.285, 0.243, 0.529, 0.022, 0.491, 0.338, 0.365, 0.295, 0.269, 0.353, 0.3, 0.425, 0.269, 0.12, 0.146, 0.454, 0.326, 0.324, 0.62, 0.205, 0.65, 0.051, 0.312, 0.327, 0.01, 0.01, 0.001, 0.017, 0.003, 0.008, 0.001, 0.001, 0.028, 0.006, 0.015, 0.054, 0.048, 0.019, 0.029, 0.107, 0.001, 0.006, 0.045, 0.02, 0.147, 0.015, 0.01, 0.01, 0.083, 0.265, 0.064, 0.553, 0.493, 0.19, 0.294, 0.416, 0.396]

5 [0.889, 0.933, 0.811, 0.998, 0.856, 0.912, 0.987, 0.942, 0.936, 0.864, 0.952, 0.706, 0.898, 0.958, 0.985, 0.942, 0.929, 0.898, 0.711, 0.988, 0.824, 0.993, 0.983, 0.983, 0.694, 0.74, 0.799, 0.857, 0.88, 0.853, 0.9, 0.932, 0.883, 0.931, 0.853, 0.843, 0.762, 0.981, 0.888, 0.919, 0.952, 0.899, 0.901, 0.909, 0.748, 0.852, 0.987, 0.931, 0.852, 0.623, 0.861, 0.908, 0.96, 0.936, 0.961, 0.935, 0.862]

5 [0.111, 0.067, 0.189, 0.002, 0.144, 0.088, 0.013, 0.058, 0.064, 0.136, 0.048, 0.294, 0.102, 0.042, 0.015, 0.058, 0.071, 0.102, 0.289, 0.012, 0.176, 0.007, 0.017, 0.017, 0.306, 0.26, 0.201, 0.143, 0.12, 0.147, 0.1, 0.068, 0.117, 0.069, 0.147, 0.157, 0.238, 0.019, 0.112, 0.081, 0.048, 0.101, 0.099, 0.091, 0.252, 0.148, 0.013, 0.069, 0.148, 0.377, 0.139, 0.092, 0.04, 0.064, 0.039, 0.065, 0.138]



Dev data

Iteration error rate

1 posi 0.6058631921824105

1. nega 0.3941368078175896

2 posi 0.6054347826086957

1. nega 0.39456521739130435

3 posi 0.6054347826086957

1. nega 0.39456521739130435

4 posi 0.6065217391304348

1. nega 0.3934782608695652

5 posi 0.6065217391304348

1. nega 0.3934782608695652

training Data

Iteration error rate

1 posi 0.6059782608695652

1. nega 0.39402173913043476

2 posi 0.6060853029068188

1. nega 0.3939146970931812

3 posi 0.6060853029068188

1. nega 0.3939146970931812

4 posi 0.6058136375984787

1. nega 0.39418636240152133

5 posi 0.6058136375984787

1. nega 0.39418636240152133

there aren’t huge differences between each iterations of trainData errors and the Dev data errors, ,meaning that the error rate is relative stable.

When the positive error rate of the training data goes up, the positive error rates of the testing data also goes up, same with the negative error rates

The fact that the training error and the Dev error are very similar also means there aren’t a lot error differences between each data sets.

1. Judging from the data, the 57 th(capital run length total) column is the majority class(its appears the most among all the columns in general)

Classifying only by the 57 th column:

false positive false negative overallError

line 1 :Fold\_ 1 0.05555555555555555 0.18457300275482094 0.10640608034744843

line 2 :Fold\_ 2 0.05206463195691203 0.19559228650137742 0.10869565217391304

line 3 :Fold\_ 3 0.05385996409335727 0.14049586776859505 0.08804347826086957

line 4 :Fold\_ 4 0.05734767025089606 0.13259668508287292 0.08695652173913043

line 5 :Fold\_ 5 0.05734767025089606 0.16298342541436464 0.09891304347826087

line 6 :Avg 0.05523509842152339 0.1632482535044062 0.09780295519992446

false positive false negative overallError

line 1 :Fold\_ 1 0.4968017057569296 0.5375626043405676 0.762214983713355

line 2 :Fold\_ 2 0.4762886597938144 0.49108589951377635 0.7782608695652173

line 3 :Fold\_ 3 0.4762886597938144 0.49108589951377635 0.7782608695652173

line 4 :Fold\_ 4 0.4732510288065844 0.4886731391585761 0.7782608695652173

line 5 :Fold\_ 5 0.4732510288065844 0.4886731391585761 0.7782608695652173All FP rata, FN rate and the overallerror rate are higher than the normal data. So overall choosing the majority class causes a higher false rate. It makes sense since the classifier was bias towards the last column and doesn’t take into account of the data from the other columns.

If the bayes data get a higher result, it means that the other columns are noises in the dataset and we actually just need the majority class as the classifier. It makes sense under the condition that judging the spam should primarily depend on the majority class only