Applied Machine Learning: Tutorial Number 3

School of Informatics, University of Edinburgh
Instructors: Oisin Mac Aodha and Siddharth N.

September 2022

1. In this question your task is to evaluate the performance of two *different* probabilistic binary classifiers, g_1 and g_2. Each classifier takes a vector of features $\boldsymbol{x}_i$ as input and estimates the probability that $\boldsymbol{x}_i$ is from the positive class i.e. $p(y_i = 1 \mid \boldsymbol{x}_i) = g_k(\boldsymbol{x}_i); k = 1, 2$. In this example, you do *not* have direct access to the classifiers, but instead only have their predictions for a set of six validation examples $\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_6\}$, along with the corresponding ground truth class labels $\{y_1, y_2, ..., y_6\}$.

i	1	2	3	4	5	6
y_i	0	1	1	0	0	1
$g_1(\boldsymbol{x}_i)$	0.1	0.9	0.7	0.3	0.2	0.2
$g_2(\boldsymbol{x}_i)$	0.2	0.2	0.6	0.7	0.6	0.3

i. Report the F-measure (i.e. F-score) for the two classifiers using their predictions above with a threshold of 0.5. Which of the two classifiers performs better on this validation set?

ii. For the second classifier only (i.e. g_2), report the false positive rate and true positive rate corresponding to the thresholds 0.0, 0.33, 0.66, and 1.0.

iii. Sketch the resulting ROC plot for g_2. You only need to show the points on the curve for the thresholds 0.0, 0.33, 0.66, and 1.0. Do you have any observations from looking at your plot?

2. The Peak Expiratory Flow Rate (PEFR) is a person's maximum speed of expiration, as measured with a peak flow meter, a small, hand-held device used to monitor a person's ability to breathe out air. It measures the airflow through the bronchi and thus the degree of obstruction in the airways.

A medical practitioner at the respiratory illness unit of a leading hospital is unhappy about having people physically blow into a flow meter due to hygiene issues and constraints on the reuse of measurement devices. They find that using scans of patients' chests and throats, they can reliably extract features that identify how wide the bronchi are, any obstructions, and how large lung capacity is. They decide to construct a regression model which they call PERP (short for Peak Expiratory Flow Predictor), that predicts PEFR using these extracted features.

Having trained PERP on some available hospital data, they want to see if their model is actually any good in practice, and so they setup an experiment with four patients that are visiting the hospital that day. They use a flow meter to measure PEFR for each patient directly, then use PERP to try and predict the same using their scans, obtaining the following data:

Patient	Flow Meter	PERP
1	15.22	17.96
2	13.74	16.17
3	22.14	23.81
4	21.66	27.28

i. Are PERP's predictions of PEFR significantly different from the flow meter's predictions? Conduct a t-test to answer this at the $\alpha = 0.05$ significance level. Use the table at the end of this document to identify the critical value c. State the null and alternative hypothesis clearly, and work through the steps required to evaluate the t-test in full.

ii. Suppose that it was found that PERP's estimation had an inadvertent error in that the intercept was estimated incorrectly, and fixing it would mean all of PERP's predictions would decrease by 0.5. Would that affect your answers from the previous part? Would the t-test's final outcome change?

3. Two students are working on a machine-learning approach to spam detection for a large bank. Each student has their *own* set of 100 labeled emails, 90% of which are used for training and 10% for validating the model. Student A runs the Naive Bayes algorithm and reports 80% accuracy on her validation set. Student B experiments with over 100 different learning algorithms, training each one on her training set, and recording the accuracy on the validation set. Her best model achieves 90% accuracy.

i. Whose algorithm would you pick for protecting a corporate network from spam and why?

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										