Machine Learning

Tutorial 5 (Error-Based Learning)

Exercise1:

You have been hired by the European Space Agency to build a model that predicts the amount of oxygen that an astronaut consumes when performing five minutes of intense physical work. The descriptive features for the model will be the age of the astronaut and their average heart rate throughout the work. The regression model is

$$\text{OXYCON} = \mathbf{w}[0] + \mathbf{w}[1] \times \text{AGE} + \mathbf{w}[2] \times \text{HEARTRATE}$$

The table that follows shows a historical dataset that has been collected for this task.

ID	OXYCON	AGE	HEART RATE
1	37.99	41	138
2	47.34	42	153
3	44.38	37	151
4	28.17	46	133
5	27.07	48	126
6	37.85	44	145
7	44.72	43	158
8	36.42	46	143
9	31.21	37	138
10	54.85	38	158
11	39.84	43	143
12	30.83	43	138

(a) Assuming that the current weights in a multivariate linear regression model are w[0] = -59.50, w[1] = -0.15, and w[2] = 0.60, make a prediction for each training instance using this model.
(b) Calculate the sum of squared errors for the set of predictions generated in Part (a).
(c) Assuming a learning rate of 0.000002, calculate the weights at the next iteration of the gradient descent algorithm.
(d) Calculate the sum of squared errors for a set of predictions generated using the new set of weights calculated in Part (c)

Exercise2:

A multivariate logistic regression model has been built to predict the propensity of shoppers to perform a repeat purchase of a free gift that they are given. The descriptive features used by the model are the age of the customer, the socioeconomic band to which the customer belongs (a, b, or c), the average amount of money the customer spends on each visit to the shop, and the average number of visits the customer makes to the shop per week. This model is being used by the marketing department to determine who should be given the free gift. The weights in the trained model are shown in the following table.

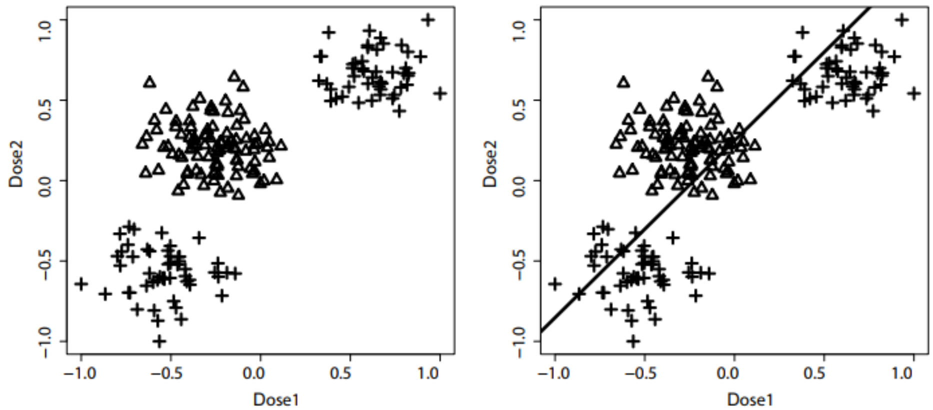
Feature	Weight
Intercept ($\mathbf{w}[0]$)	-3.82398
AGE	-0.02990
SOCIOECONOMIC BAND B	-0.09089
SOCIOECONOMIC BAND C	-0.19558
SHOP VALUE	0.02999
SHOP FREQUENCY	0.74572

Use this model to make predictions for each of the following query instances.

ID	AGE	SOCIOECONOMIC BAND	SHOP FREQUENCY	SHOP VALUE
1	56	b	1.60	109.32
2	21	c	4.92	11.28
3	48	b	1.21	161.19
4	37	c	0.72	170.65
5	32	a	1.08	165.39

Exercise3:

The effects that can occur when different drugs are taken together can be difficult for doctors to predict. Machine learning models can be built to help predict optimal dosages of drugs so as to achieve a medical practitioner's goals. In the following figure, the image on the left shows a scatter plot of a dataset used to train a model to distinguish between dosages of two drugs that cause a dangerous interaction and those that cause a safe interaction. There are just two continuous features in this dataset, DOSE1 and DOSE2 (both normalized to the range (-1,1) using range normalization), and two target levels, dangerous and safe. In the scatter plot, DOSE1 is shown on the horizontal axis, DOSE2 is shown on the vertical axis, and the shapes of the points represent the target level-crosses represent dangerous interactions and triangles represent safe interactions.



In the preceding figure, the image on the right shows a simple linear logistic regression model trained to perform this task. This model is

$$P(\text{TYPE} = \textit{dangerous}) = \\ \textit{Logistic}(0.6168 + 2.7320 \times \text{DOSE1} - 2.4809 \times \text{DOSE2})$$

Plainly, this model is not performing well.

(a) Would the similarity-based, information-based, or probability-based predictive modeling approaches already covered be likely to do a better job of learning this model than the simple linear regression model?

(b) A simple approach to adapting a logistic regression model to learn this type of decision boundary is to introduce a set of basis functions that will allow a nonlinear decision boundary to be learned. In this case, a set of basis functions that generate a cubic decision boundary will work well. An appropriate set of basis functions is as follows:

$$\begin{aligned}
&\phi_0(\langle \text{DOSE1}, \text{DOSE2} \rangle) = 1 && \phi_1(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE1} \\
&\phi_2(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE2} && \phi_3(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE1}^2 \\
&\phi_4(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE2}^2 && \phi_5(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE1}^3 \\
&\phi_6(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE2}^3 && \phi_7(\langle \text{DOSE1}, \text{DOSE2} \rangle) = \text{DOSE1} \times \text{DOSE2}
\end{aligned}$$

Training a logistic regression model using this set of basis functions leads to the following model:

$$\begin{aligned}
&P(\text{TYPE} = \textit{dangerous}) = \\
&\textit{Logistic}\big(-0.848 \times \phi_0(\langle \text{DOSE1}, \text{DOSE2} \rangle) + 1.545 \times \phi_1(\langle \text{DOSE1}, \text{DOSE2} \rangle) \\
&\quad - 1.942 \times \phi_2(\langle \text{DOSE1}, \text{DOSE2} \rangle) + 1.973 \times \phi_3(\langle \text{DOSE1}, \text{DOSE2} \rangle) \\
&\quad + 2.495 \times \phi_4(\langle \text{DOSE1}, \text{DOSE2} \rangle) + 0.104 \times \phi_5(\langle \text{DOSE1}, \text{DOSE2} \rangle) \\
&\quad + 0.095 \times \phi_6(\langle \text{DOSE1}, \text{DOSE2} \rangle) + 3.009 \times \phi_7(\langle \text{DOSE1}, \text{DOSE2} \rangle)\big)
\end{aligned}$$

Use this model to make predictions for the following query instances:

ID	DOSE1	DOSE2
1	0.50	0.75
2	0.10	0.75
3	-0.47	-0.39
4	-0.47	0.18

Exercise 4:

The following **multinomial logistic regression** model predicts the TYPE of a retail customer (*single*, *family*, or *business*) on the basis of the average amount that they spend per visit, SPEND, and the average frequency of their visits, FREQ:

$$\begin{aligned}
\mathbb{M}_{\mathbf{w}_{single}}(\mathbf{q}) &= \textit{logistic}(0.7993 - 15.9030 \times \text{SPEND} + 9.5974 \times \text{FREQ}) \\
\mathbb{M}_{\mathbf{w}_{family}}(\mathbf{q}) &= \textit{logistic}(3.6526 + -0.5809 \times \text{SPEND} - 17.5886 \times \text{FREQ}) \\
\mathbb{M}_{\mathbf{w}_{business}}(\mathbf{q}) &= \textit{logistic}(4.6419 + 14.9401 \times \text{SPEND} - 6.9457 \times \text{FREQ})
\end{aligned}$$

Use this model to make predictions for the following query instances:

ID	SPEND	FREQ
1	-0.62	0.10
2	-0.43	-0.71
3	0.00	0.00

Exercise 5:

A support vector machine has been built to predict whether a patient is at risk of cardiovascular disease. In the dataset used to train the model, there are two target levels-*high risk* (the positive level +1) or *low risk* (the negative level -1) and three descriptive features: AGE, BMI, and BLOOD PRESSURE. The support vectors in the trained model are shown in the table below (all descriptive feature values have been standardized).

AGE	BMI	BLOOD PRESSURE	RISK
-0.4549	0.0095	0.2203	low risk
-0.2843	-0.5253	0.3668	low risk
0.3729	0.0904	-1.0836	high risk
0.558	0.2217	0.2115	high risk

In the model the value of w_0 is -0.0216, and the values of the α parameters are <1.6811, 0.2384, 0.2055, 1.7139> What predictions would this model make for the following query instances?

ID	AGE	BMI	BLOOD PRESSURE
1	-0.8945	-0.3459	0.5520
2	0.4571	0.4932	-0.4768
3	-0.3825	-0.6653	0.2855
4	0.7458	0.1253	-0.7986

Exercise 6:

Consider the training dataset for binary classification, consisting of the features x, y and class labels:

ID	x	y	Class
1	1	2	+
2	2	1	-
3	1	3	+
4	3	1	-

1) Plot the dataset along with the optimal separator line Δ, along with Δ_+ and Δ_-, the two lines parallel to Δ, closest to the positive and negative examples respectively.
2) Given that the equations of $(\Delta, \Delta_+, \Delta_-)$ are:
$$\Delta: w_1 x + w_2 y + b = 0$$
$\Delta_+: w_1 x + w_2 y + b = +a$
$\Delta_-: w_1 x + w_2 y + b = -a.$
Express the margin between the Δ_+ and Δ_- in terms of the line parameters w_1, w_2 and b and the positive offset a.
3) Prove that scaling the equations of $(\Delta, \Delta_+, \Delta_-)$ by a nonzero positive factor k does not change the separating lines or margin.
4) Examine the equations representing the lines $(\Delta, \Delta_+, \Delta_-)$. Does setting the offset $a = 1$ result in any loss of generality? Explain.
5) To determine the maximum margin separator, present the quadratic function that needs to be minimized along with all the constraints on the parameters (w_1, w_2, b).
6) Assuming the bias term b is set to zero, sketch the feasible region in the (w_1, w_2) space satisfying the constraints of the previous question.
7) Visually estimate the values of (w_1, w_2) that maximize the margin between the two classes.
8) Using the feasible region, explain graphically why certain data points do not contribute to determining the maximum margin separator.