Data Analysis In-Class Worksheet #14: Logistic Regression

Student Name _____Section ⬜1 ⬜ 2

Part #1: Logistic Regression Fundamentals

What does the "odds" measure? _____

If the probability of winning a 30% coupon at Kohl's is 10%, then what's the odds of winning a 30% coupon? _____

$\log_{10}(10000)$ = ____

Why do we need to use log(odds), not probabilities, in estimating logistic regression?

Reason #1: _____

Reason #2: _____

Suppose we have this logistic regression model.

$\hat{Y} = 2 + 1 * BED$

If BED = 3, then $\hat{Y}$ = _____

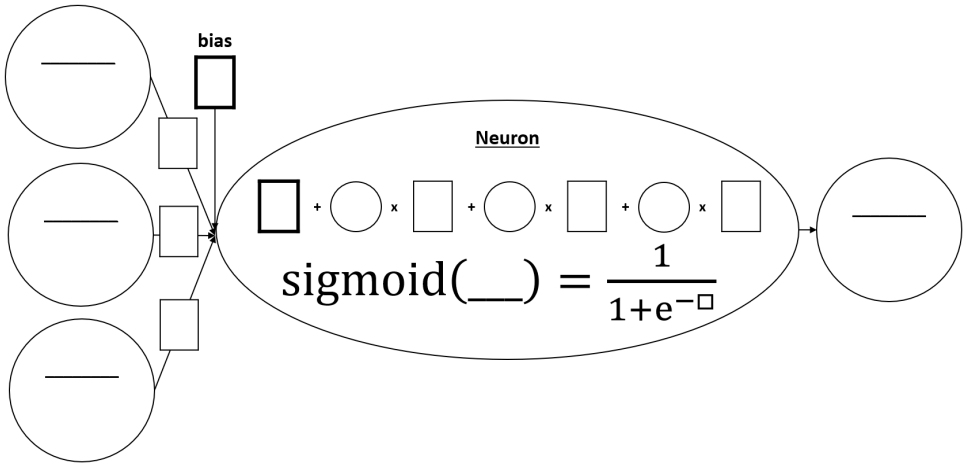
$\hat{Y}$ is ⬜ probability ⬜ odds ⬜ ln(odds)

Convert $\hat{Y}$ back into probability using the _____ function = ———————— = _____
(Reminder: You can use the formula set up in Google Sheets to compute the value.)

Why do we need to convert $\hat{Y}$ back into probability?
_____

Make a binary decision using the probability:
Convert any probability > .5 into YES, and any probability <= .5 into NO. The decision is _____.

Illustrate this logistic regression model as a deep learning neuron with a sigmoid unit:



Part #2: Logistic Regression Application

Does Property Age predict the Property's Zipcode Location?

To answer this question, we can rephrase it as a pair of hypotheses:

H_0: _____

H_1: _____

Express these hypotheses in terms of the regression β_1 slope

H_0: _____

H_1: _____

For H_0: $\hat{Y}$ = _____

For H_1: $\hat{Y}$ = _____

In our class dataset, Zipcode has only two possible outcomes: 23185 and 23188. Therefore, Zipcode is a _____ variable.

- ☐ Binary (categorical)
- ☐ Numeric

If the probability (p) of being in 23185 is 80%, then the probability of being in 23188 would be ____%. The odds of being in 23185 would be ____ / ___ = _____. The natural log of this odds is _____

If the probability (p) of being in 23185 is 50%, then the probability of being in 23188 would be ____% The odds of being in 23185 would be ____ / ___ = _____. The natural log of this odds (ln(odds)) is _____

Coefficients

				Wald Test		
	Estimate	Standard Error	z	Wald Statistic	df	p
(Intercept)	0.152	0.327	0.466	0.217	1	0.641
Age	−0.029	0.011	−2.532	6.412	1	0.011

Note. Zipcode level '23188' coded as class 1.

$\hat{Y}$ = _____ + _____ * Age

The predicted value ($\hat{Y}$) from a logistic regression is

⬜ Probabilities

⬜ Binary categories

⬜ Odds

⬜ Natural log of odds (ln(odds) or logit)

The intercept is _____, which tells us _____

The slope is ____, which tells us _____

When a property is a year older, its odds of being in 23188 (Class 1) would _____

Abdullah's $\hat{Y}$ = 0.429. Convert it into odds: _____

Convert odds into probability value: _____

Use the probability value to predict if the property is in 23185 or 23188:

_____

Abdullah's property is located in 23188.
This is a ⬜ correct ⬜ incorrect prediction.

If we get a confusion matrix like this one, calculate the overall accuracy rate:

_____

		Predicted	
		23185	23188
Actual	23185	4	6
	23188	16	24

APA Writeup

A logistic regression was performed to ascertain the effect of _____ on the likelihood of _____. The logistic regression model was statistically (in)significant, x^2 (____) = _____, p ____, McFadden's R^2 = _____. The model correctly classified _____% of cases.