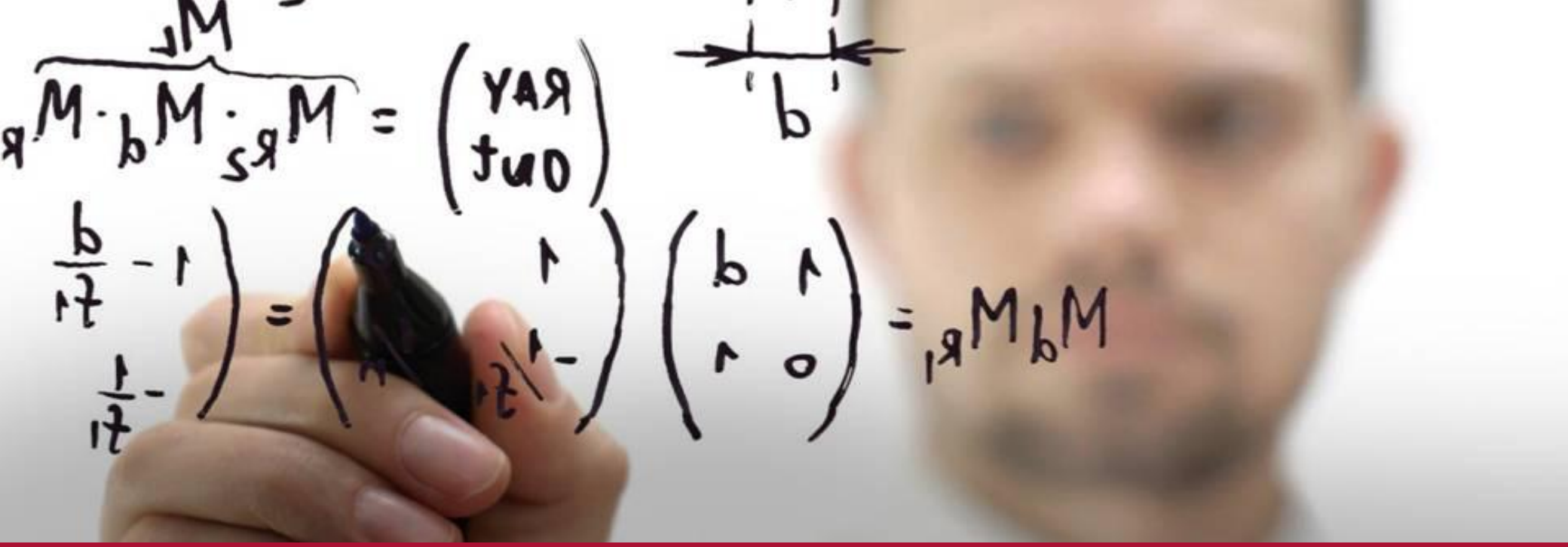


Lesson X : Regression Algorithms

Model Essentials: Regressions

- **Predict new cases.**
- **Select useful inputs.**
- **Optimize complexity.**

Prediction formula

Sequential selection

Best model from sequence

Linear Regression Prediction Formula

$$\hat{y} = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

- $\hat{y}$: prediction estimate
- $\hat{w}_0$: intercept estimate
- $\hat{w}_1$: parameter estimate
- x_1: input measurement
- $\hat{w}_2$: parameter estimate
- x_2: input measurement

Choose intercept and parameter estimates to *minimize*:

$$\sum_{\text{training data}} (y_i - \hat{y}_i)^2$$

- $\sum$: squared error function
- y_i: training data
- $\hat{y}_i$: prediction estimate

Linear Regression Prediction Formula

$$\hat{y} = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

- $\hat{y}$: prediction estimate
- $\hat{w}_0$: intercept estimate
- $\hat{w}_1$, $\hat{w}_2$: parameter estimates
- x_1, x_2: input measurement

Choose intercept and parameter estimates to *minimize*.

squared error function

$$\sum_{\text{training data}} (y_i - \hat{y}_i)^2$$

Logistic Regression Prediction Formula

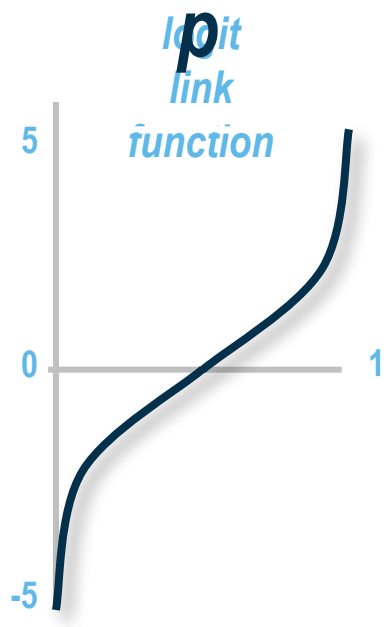
$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

logit scores

Logit Link Function

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

logit scores

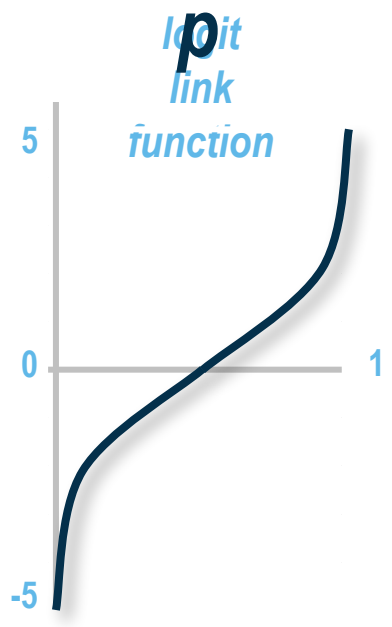


The logit link function transforms probabilities (between 0 and 1) to logit scores (between −∞ and +∞).

Logit Link Function

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

logit scores



The logit link function transforms probabilities (between 0 and 1) to logit scores (between −∞ and +∞).

Logit Link Function

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2 = \text{logit}(\hat{p})$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

To obtain prediction estimates, the logit equation is solved for p.

Logit Link Function

$$\log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2 = \text{logit}(\hat{p})$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

To obtain prediction estimates, the logit equation is solved for p.

Logit Link Function

$$\text{logit}(\hat{p}) = \hat{w}_0 + \hat{w}_1 \cdot x_1 + \hat{w}_2 \cdot x_2$$

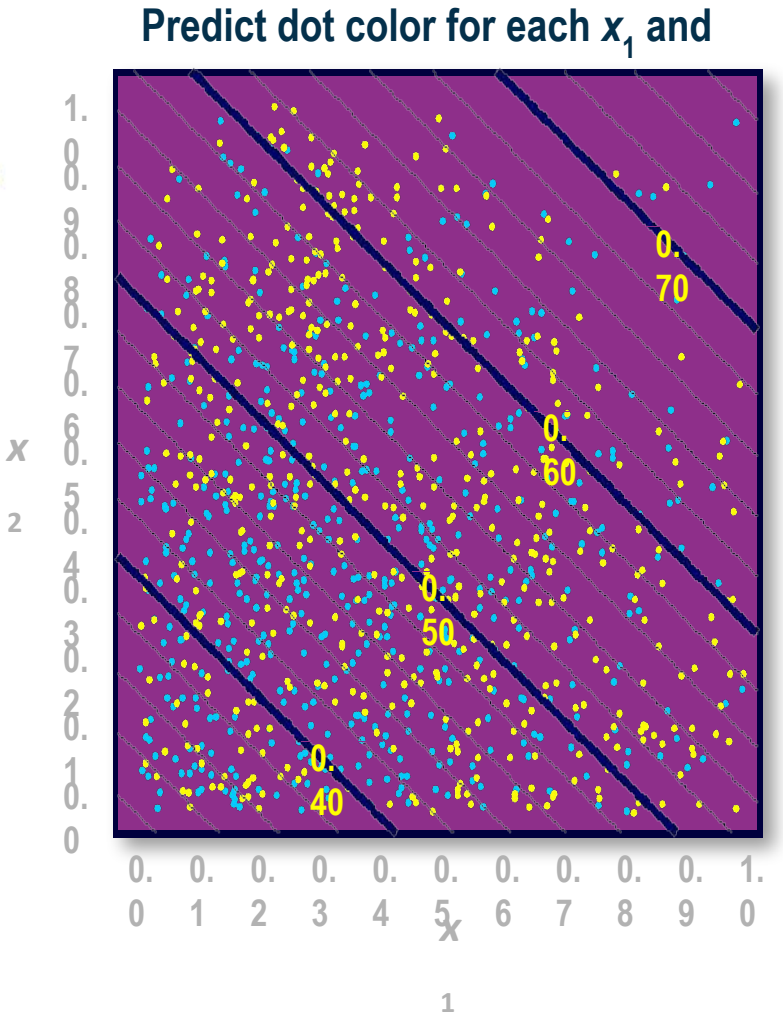
$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

Simple Prediction Illustration: Regressions

$$\text{logit}(\hat{p}) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

You need intercept and parameter estimates.

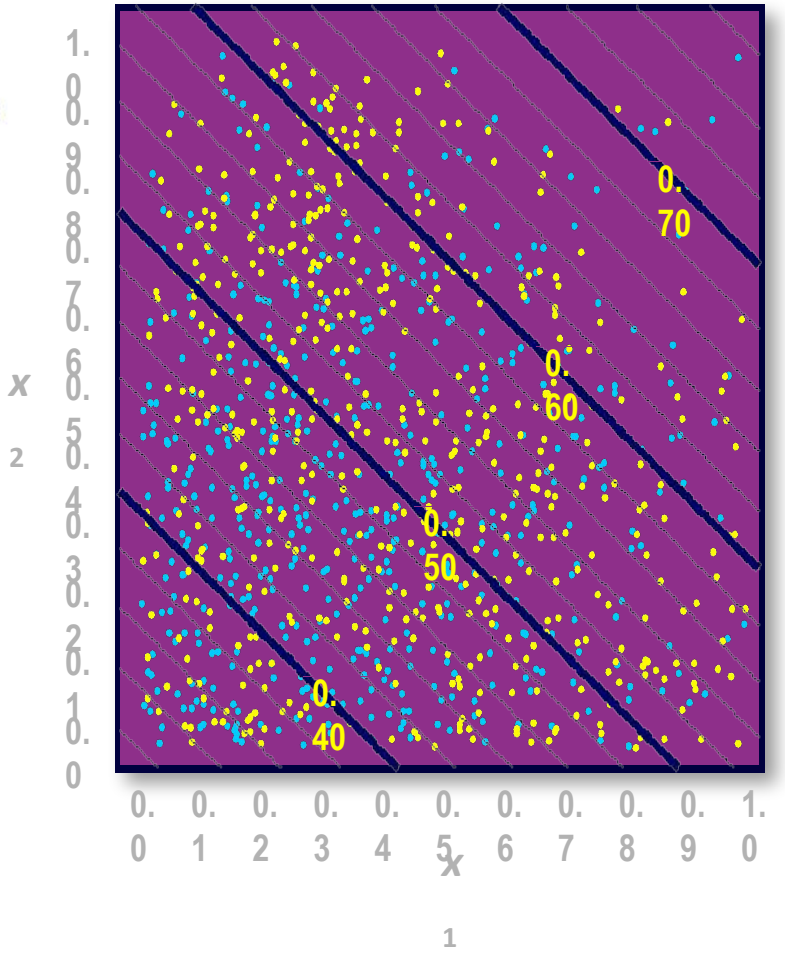


Simple Prediction Illustration: Regressions

$$\text{logit}(\hat{p}) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

You need intercept and parameter estimates.



Simple Prediction Illustration: Regressions

$$\text{logit}(\hat{p}) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

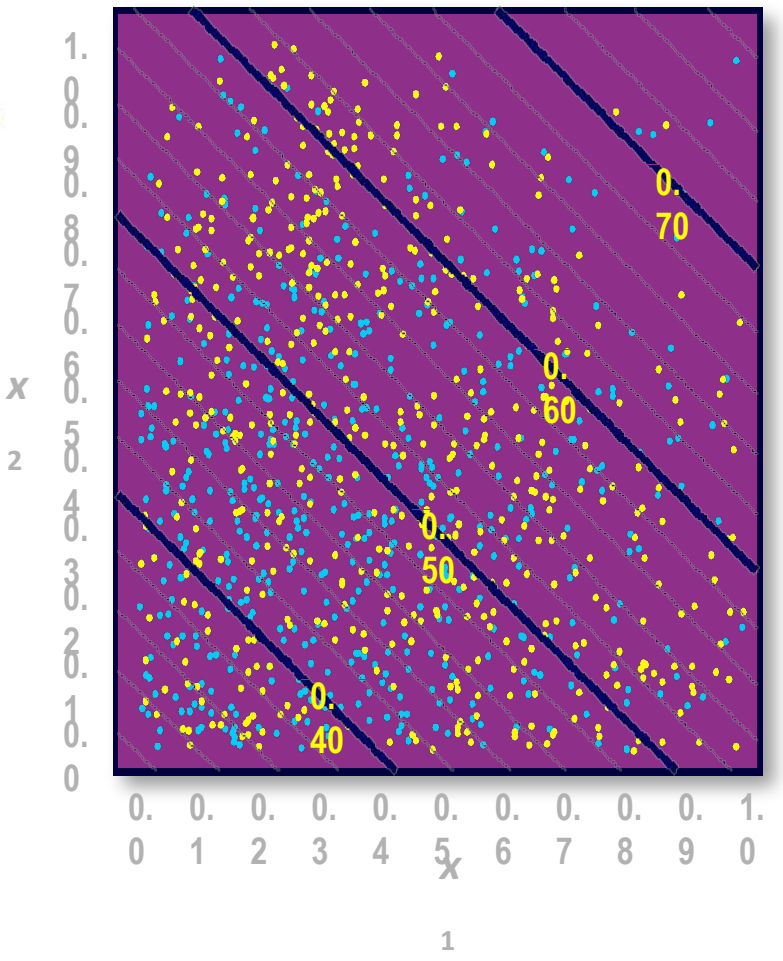
$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

Find parameter estimates by *maximizing*

$$\sum \log(\hat{p}_i) + \sum \log(1 - \hat{p}_i)$$

primary outcome training cases — *secondary outcome training cases*

log-likelihood function



Simple Prediction Illustration: Regressions

$$\text{logit}(\hat{p}) = \hat{w}_0 + \hat{w}_1 x_1 + \hat{w}_2 x_2$$

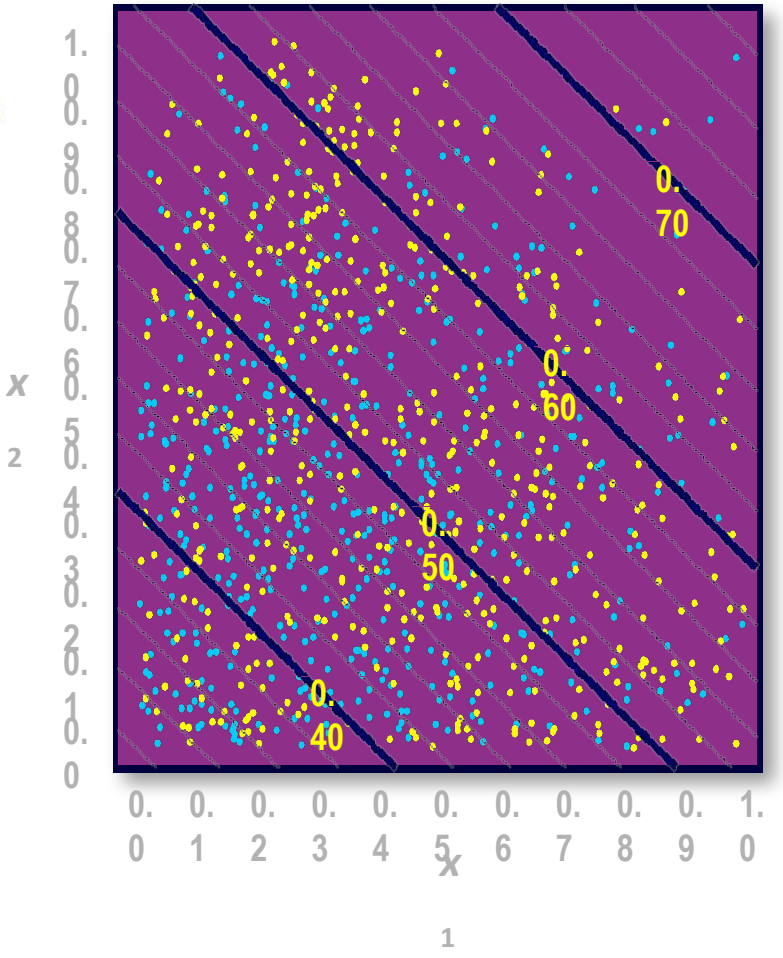
$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

Find parameter estimates by *maximizing*

$$\sum \log(\hat{p}_i) + \sum \log(1 - \hat{p}_i)$$

primary outcome training cases — *secondary outcome training cases*

log-likelihood function

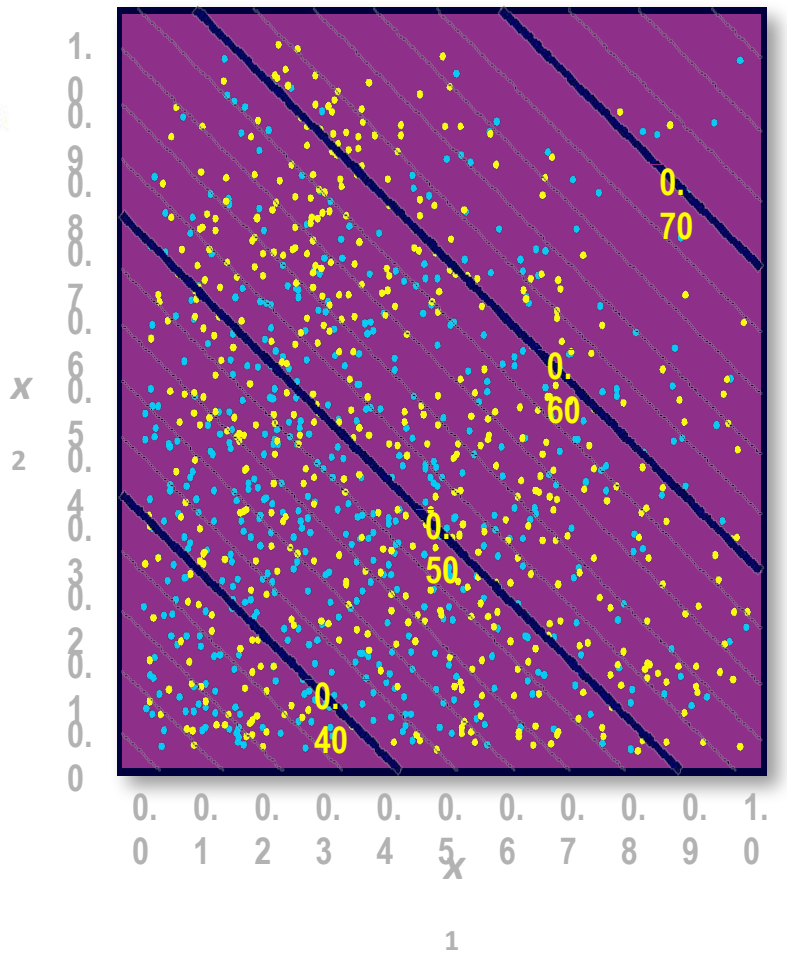


Simple Prediction Illustration: Regressions

$$\text{logit}(\hat{p}) = -0.81 + 0.92\,x_1 + 1.11\,x_2$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$

Using the maximum likelihood estimates, the prediction formula assigns a logit score to each x_1 and x_2.



4.01 Multiple Choice Question

What is the logistic regression prediction for the indicated point?

a. 0.243

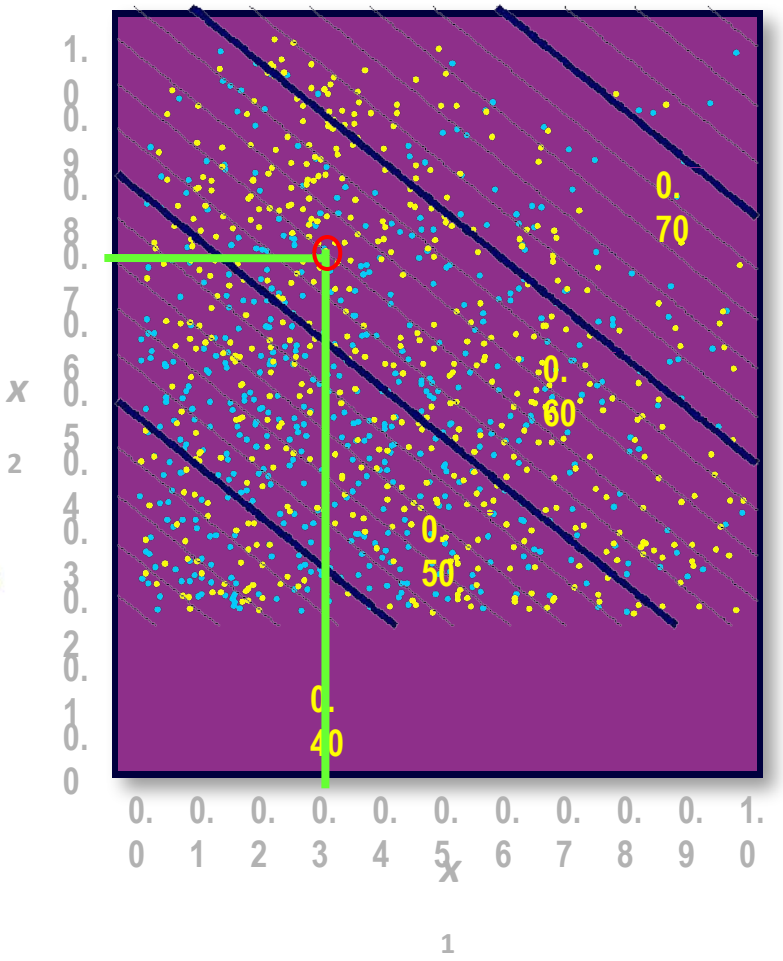
b. 0.56

c. yellow

d. It depends.

$$\text{logit}(\hat{p}) = -0.81 + 0.92\,x_1 + 1.11\,x_2$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$



4.01 Multiple Choice Question – Correct Answer

What is the logistic regression prediction for the indicated point?

a. 0.243

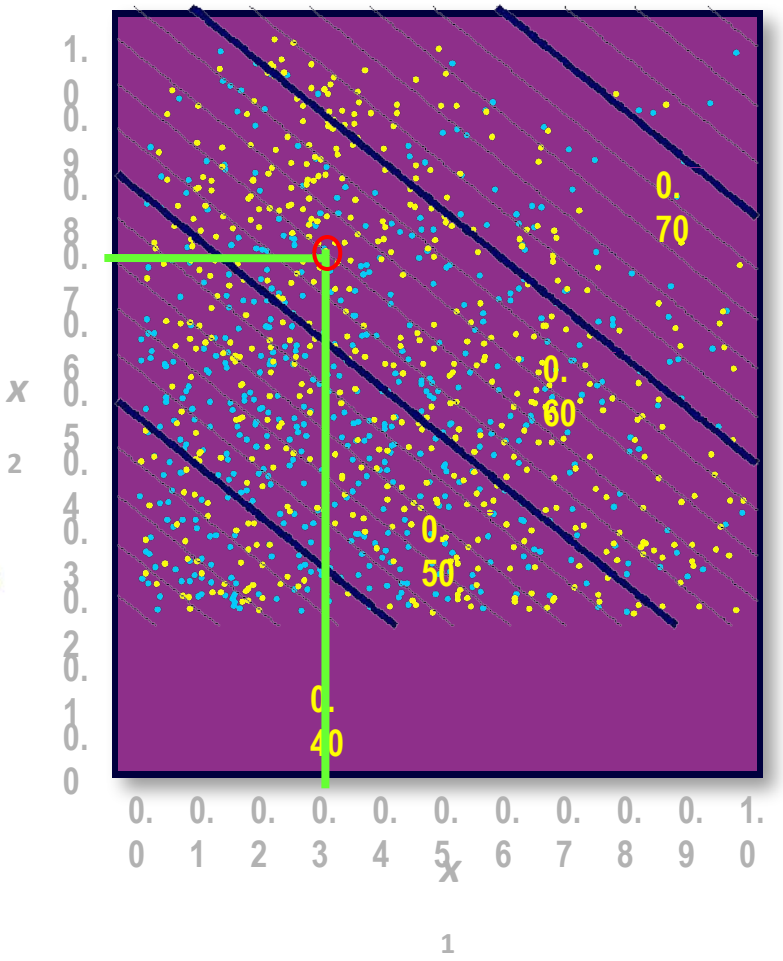
b. 0.56

c. yellow

d. It depends.

$$\text{logit}(\hat{p}) = -0.81 + 0.92\,x_1 + 1.11\,x_2$$

$$\hat{p} = \frac{1}{1 + e^{-\text{logit}(\hat{p})}}$$



Regressions: Beyond the Prediction Formula

- **Manage missing values.**
- **Interpret the model.**
- **Handle extreme or unusual values.**
- **Use nonnumeric inputs.**
- **Account for nonlinearities.**

Construct a Regression using SKlearn

ADSUP Lesson VIII Ra

1- Construct a Regression model for Fitness Dataset

Practice

This practice reinforces the concepts discussed previously.

INSTITUTE OF DATA SCIENCE

Science for real life

IDS Analytics is a firm of experts specialized in applying quantitative methods to growth and optimization issues of enterprises, particularly in risk, revenue management and marketing domains.

We help enterprises both substantially improve commercial success and build capacity to repeat this performance continuously. In true "scientific" spirit, IDS analytics provides a comprehensive and complementing range of services across management advisory, capability building and managed services disciplines.

www.ids-analytics.com

OFİS
Telefon +90 (216) 343 6623
Tophanelioğlu Cad. Korukent Köşk
No:6/5 Koşuyolu
İstanbul TÜRKİYE