greatlearning
Learning for Life

Machine Learning

Logistic Regression

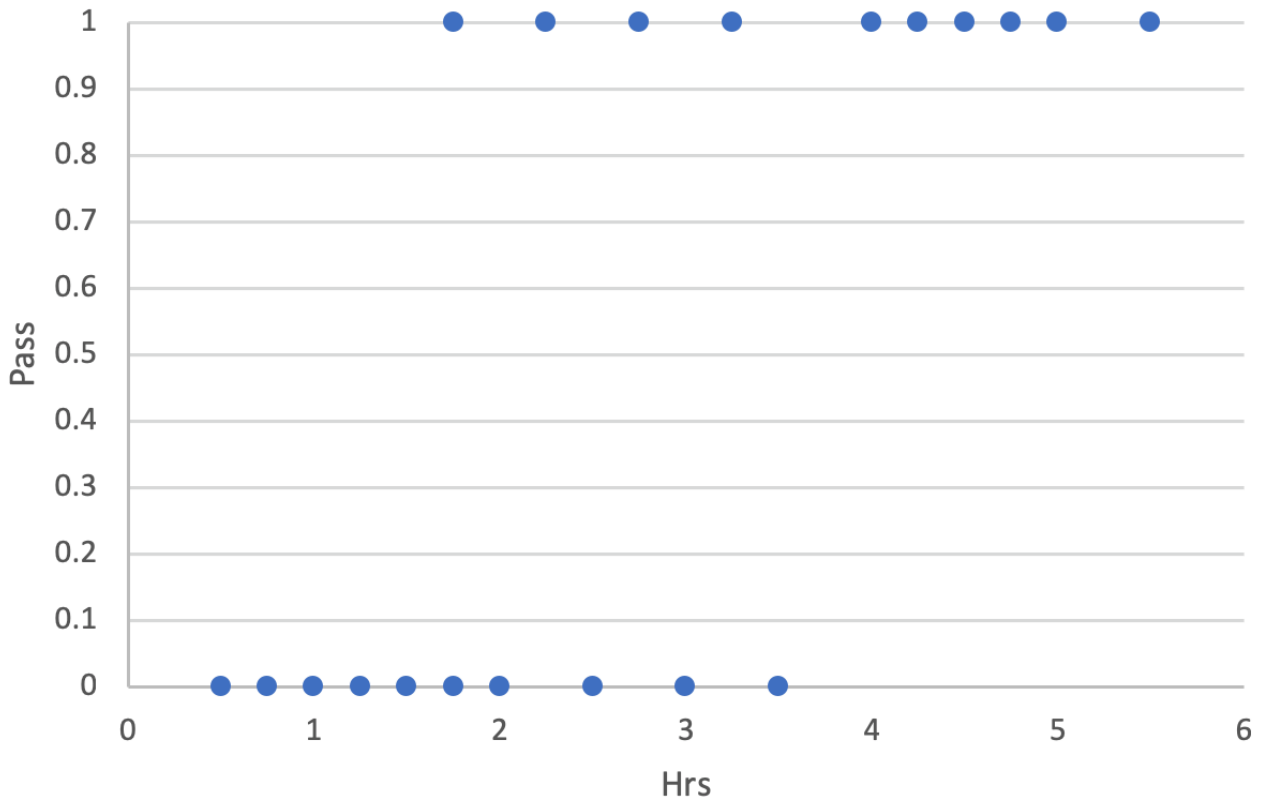
Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Linear Regression for Classification?

greatlearning
Learning for Life

- How likely is a student to pass if he/she studies for 5 hrs?
  - Using data from 20 students
- Classification problem! Can use linear regression?

Hrs	Pass?
0.5	0
0.75	0
1	0
1.25	0
1.5	0
1.75	0
1.75	1
2	0
2.25	1
2.5	0
2.75	1
3	0
3.25	1
3.5	0
4	1
4.25	1
4.5	1
4.75	1
5	1
5.5	1



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Source: Wikipedia

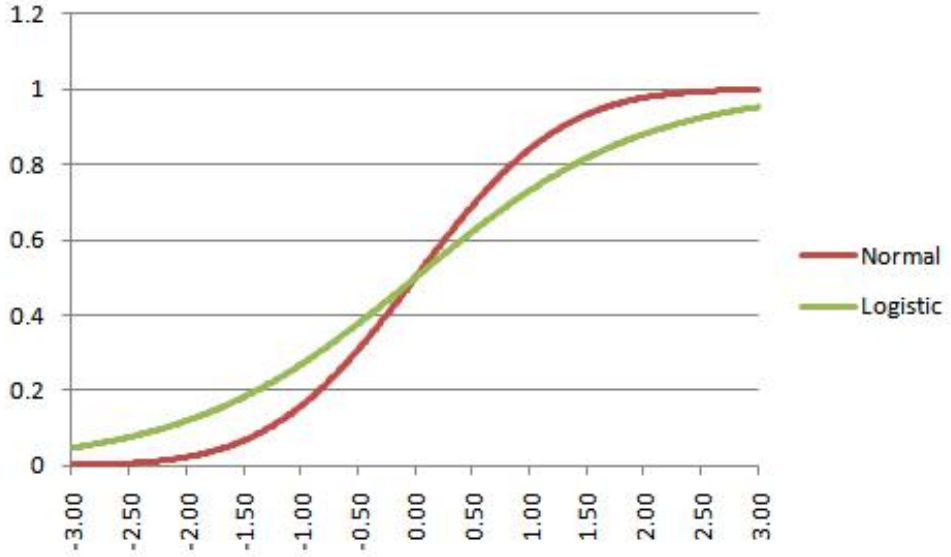
Instead can we fit a curve?

greatlearning
Learning for Life

- Regression fits $y = a + bx$
- Instead why not fit?

$$y = f(a + bx)$$

- Common choices for $f()$
  - Logistic Regression: $y = \dfrac{1}{1 + e^{-(a+bx)}}$
  - Probit Regression: $y = \Phi(x)$



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

The Logit function

- Logit function: $y = \frac{1}{1 + e^{-(a+bx)}}$
- Equivalent to thinking of $\log\left(\frac{y}{1-y}\right) = a + bx$

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Finding the best fit logic curve?

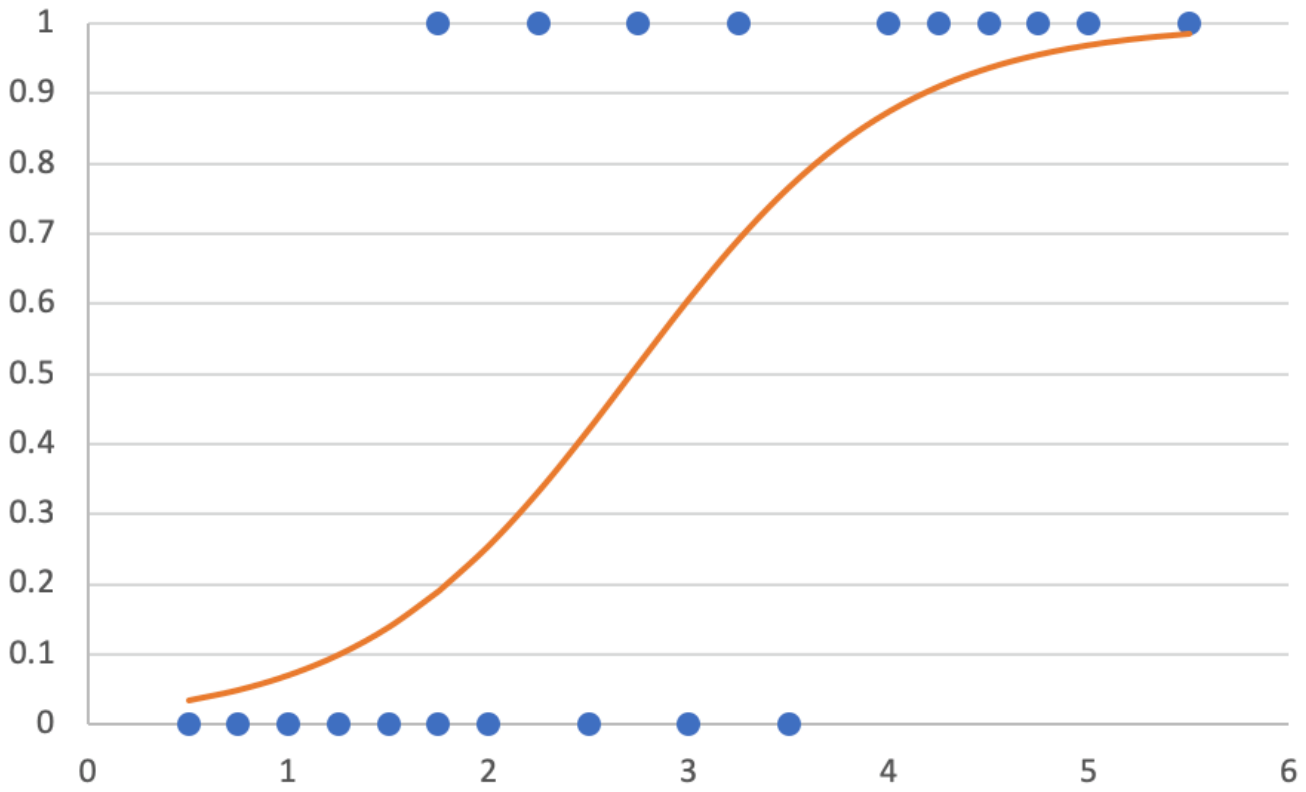
greatlearning
Learning for Life

- Linear regression minimized sum of squared residuals. This unfortunately will not work in logistic regression!
- Instead we choose to minimize the “Log Loss” or “Cross-Entropy”

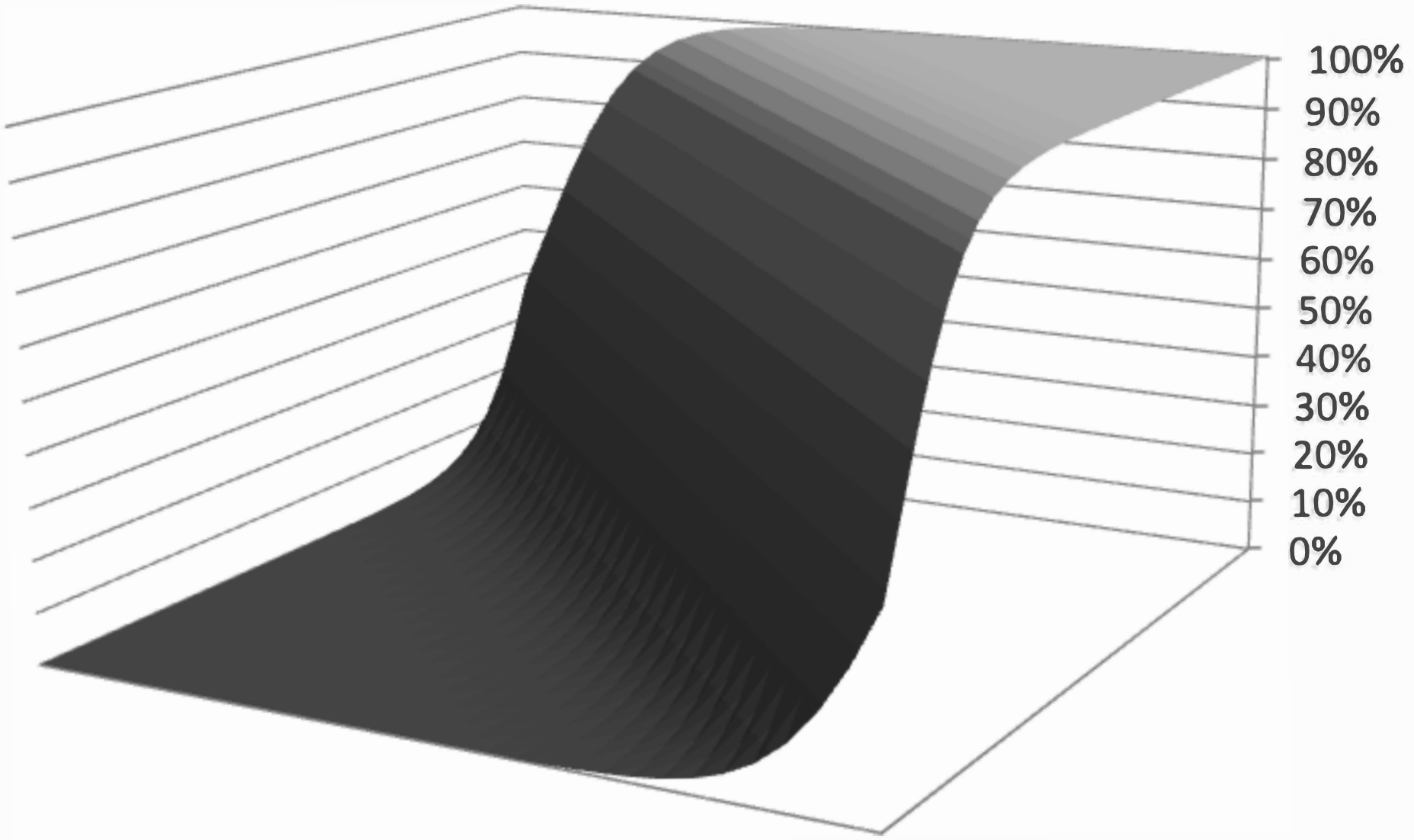
$$-y\log(\hat{y}) - (1-y)\log(1-\hat{y})$$

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

- How likely is a student to pass if he/she studies for 5 hrs?



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited



Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Logistic Reg - Pros and Cons

greatlearning
Learning for Life

- Advantages
  - A classification model that does give probabilities
  - Easily extended to multiple classes (multinomial regression)
  - Quick to train and very fast at classifying unknown records
- Disadvantages
  - Constructs linear boundaries
  - Assumes that variables are independent (eg. does not include interaction terms)
  - Interpretation of coefficients is difficult

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited