Model Performance: Accuracy and Cross-Entropy Loss

Accuracy

- Defined as: $\frac{\text{correct predictions}}{\text{total predictions}}$

- What is a correct prediction?
  - Recall that in logistic regression, our predicted value is $p(y_i = 1)$, the probability of event occurrence
  - To decide whether the prediction is correct, we must convert $p(y_i = 1)$ to a binary prediction by setting a *threshold*

$p(y_i = 1)$	y_i	**Correct?**
.77	1	
.27	0	
.53	0	
.54	1	
.11	0	
.87	0	

Accuracy

- Easy to understand, but not useful in learning
- Rarely used as a performance metric; we will discuss performance metrics in subsequent lectures

Cross-Entropy Loss

- A *loss* is a measure of model performance that we use during training
- The lower the loss, the better the performance. During training, we look for parameters that *minimize* the loss.
  - Sometimes, including in logistic regression, we can be sure that we’ve found the best possible parameters – those that give us the lowest loss possible
  - Other times we cannot be sure, so we lower the loss as much as we can

Cross-Entropy Loss

- Suppose we have a fair coin:
  - $p(y_i = 1) = .5$ (heads)
  - $p(y_i = 0) = .5$ (tails)

Cross-Entropy Loss

- What is the probability that we observe the following?

.5 x .5 x .5

Cross-Entropy Loss

- What if the coin is not balanced?
  - $p(y_i = 1) = .3$ (heads)
  - $p(y_i = 0) = .7$ (tails)

Cross-Entropy Loss

- What is the probability that we observe the following?



.7 x .3 x .7

Cross-Entropy Loss

- Suppose we don't know $p(y_i = 1)$. We might try to infer it – i.e. *learn* it – by choosing a value that maximizes the probability of our observations. We may need many observations to be confident.



? x ? x ?

Cross-Entropy Loss

- What if it’s a different coin every time…

.5 x .3 x .6

Cross-Entropy Loss

- …or not a coin

$y_1 = 1$ (dies) $\qquad$ $y_2 = 0$ (survives) $\qquad$ $y_3 = 0$ (survives)

p_1? $\quad$ x $\quad$ p_2? $\quad$ x $\quad$ p_3?

Cross-Entropy Loss

- We'd like to predict p_i values that maximize the probability of our observations (the y_i). This is what the cross-entropy loss measures.

$y_1 = 1$ (dies)

p_1?

x

$y_2 = 0$ (survives)

p_2?

x

$y_3 = 0$ (survives)

p_3?

Cross-Entropy Loss

- We'd like to predict p_i values that maximize the probability of our observations (the y_i). This is what the cross-entropy loss measures.

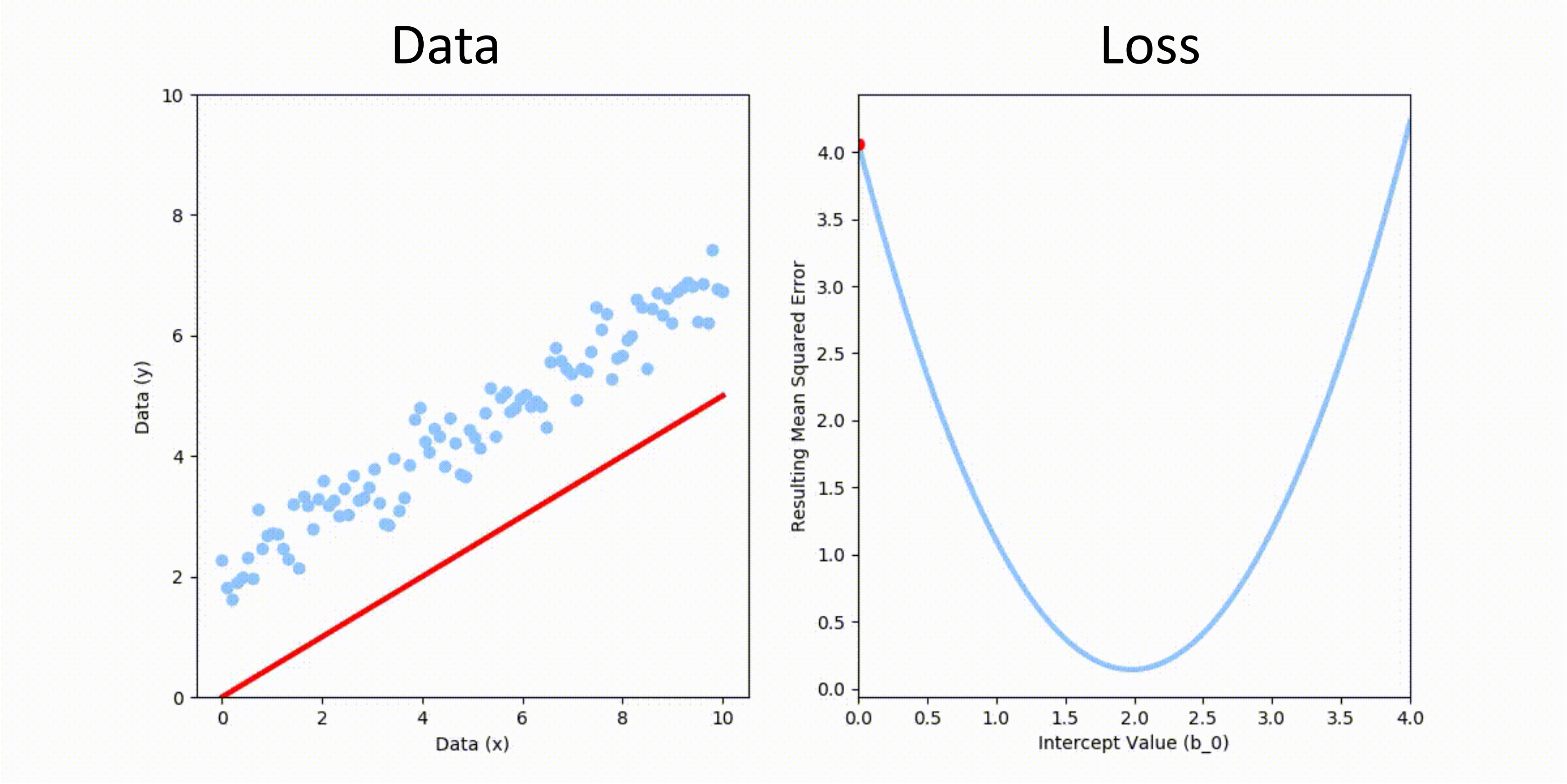
- For those who are interested: the cross-entropy loss is the *negative log-likelihood* of our observations, given our predictions.

- If we reduce it, we increase the probability of our observations.

- If we increase it, we reduce the probability of our observations.

Minimizing Loss

- As we change a parameter – in this case, the *intercept*, or *bias* of a linear regression model – the loss changes.



Minimizing Loss

- How do we reduce the loss? Follow the slope.