LOGISTIC REGRESSION
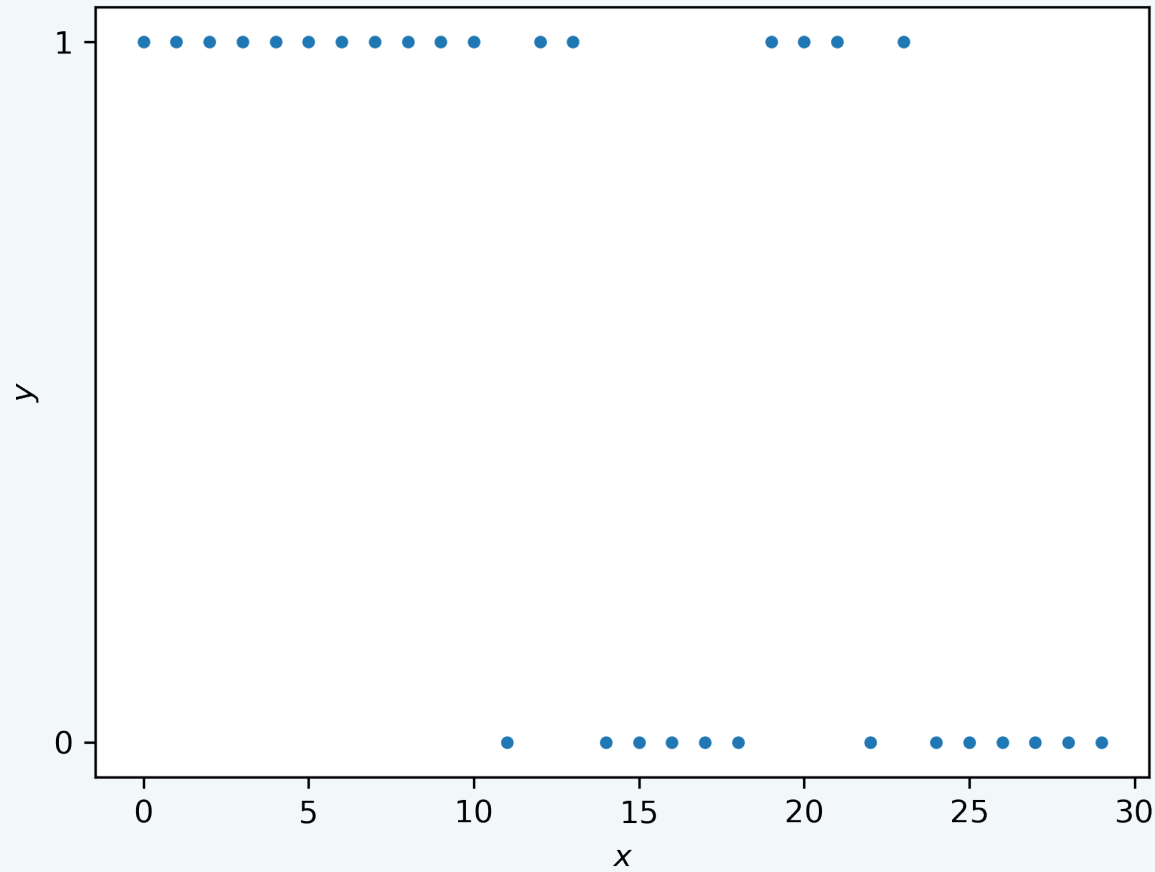
Objectives

▶ Recognize similarities between linear and logistic regression

▶ Understand the role of the logit and inverse logit in logistic regression
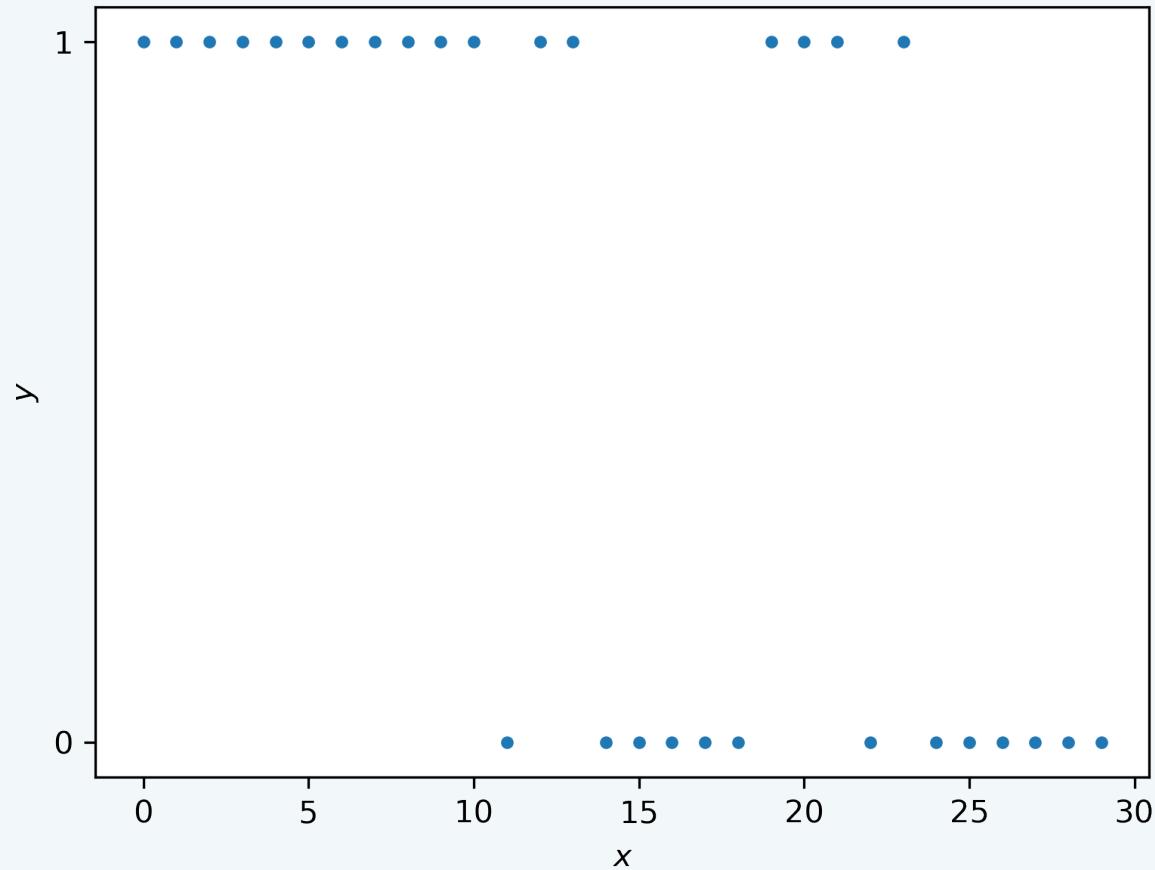
Example: Shooting Baskets

- Let's try shooting baskets from lots of different distances
- What effect would you expect? What do you see?

Example: Shooting Baskets

- Let's try shooting baskets from lots of different distances
- What effect would you expect? What do you see?
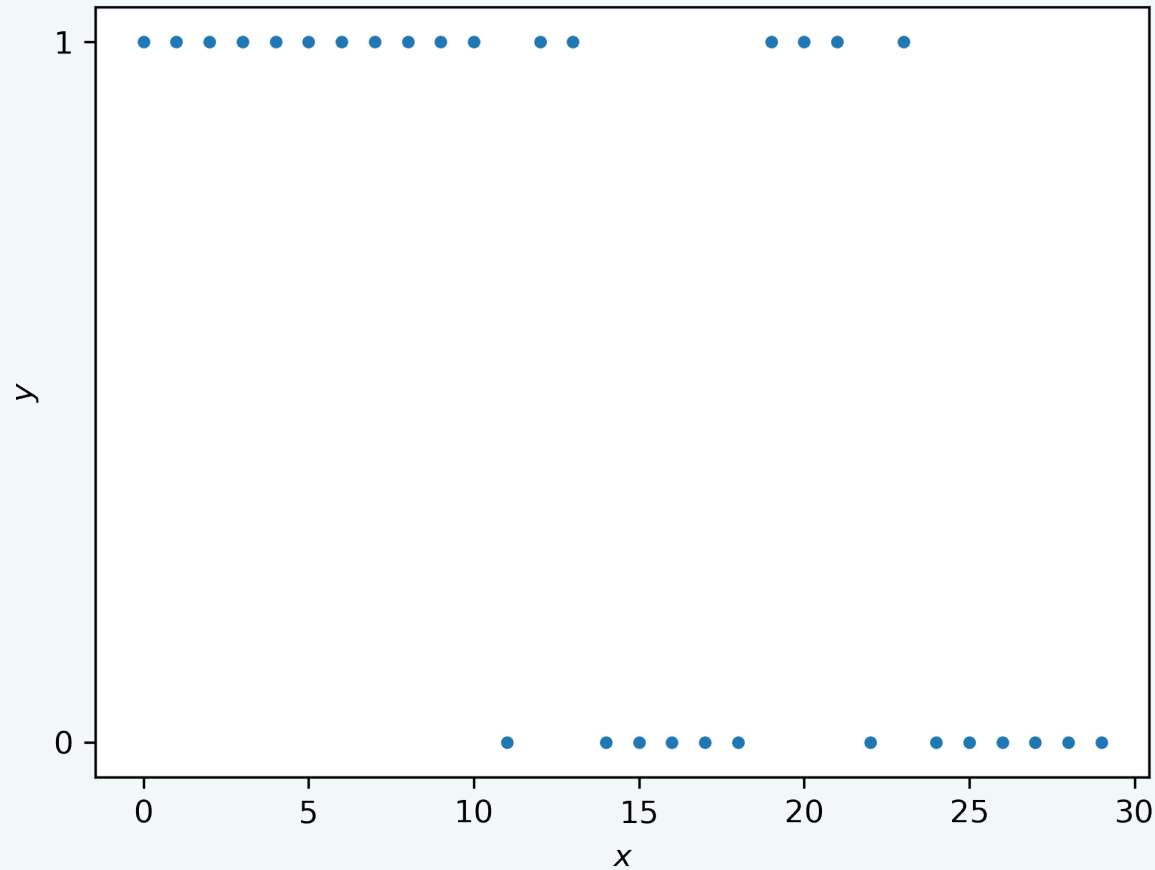- Can we apply the "transformation" approach here?

Example: Shooting Baskets

- Let's try shooting baskets from lots of different distances
- What effect would you expect? What do you see?
- Can we apply the "transformation" approach here?
- The two y values would map into two different y values
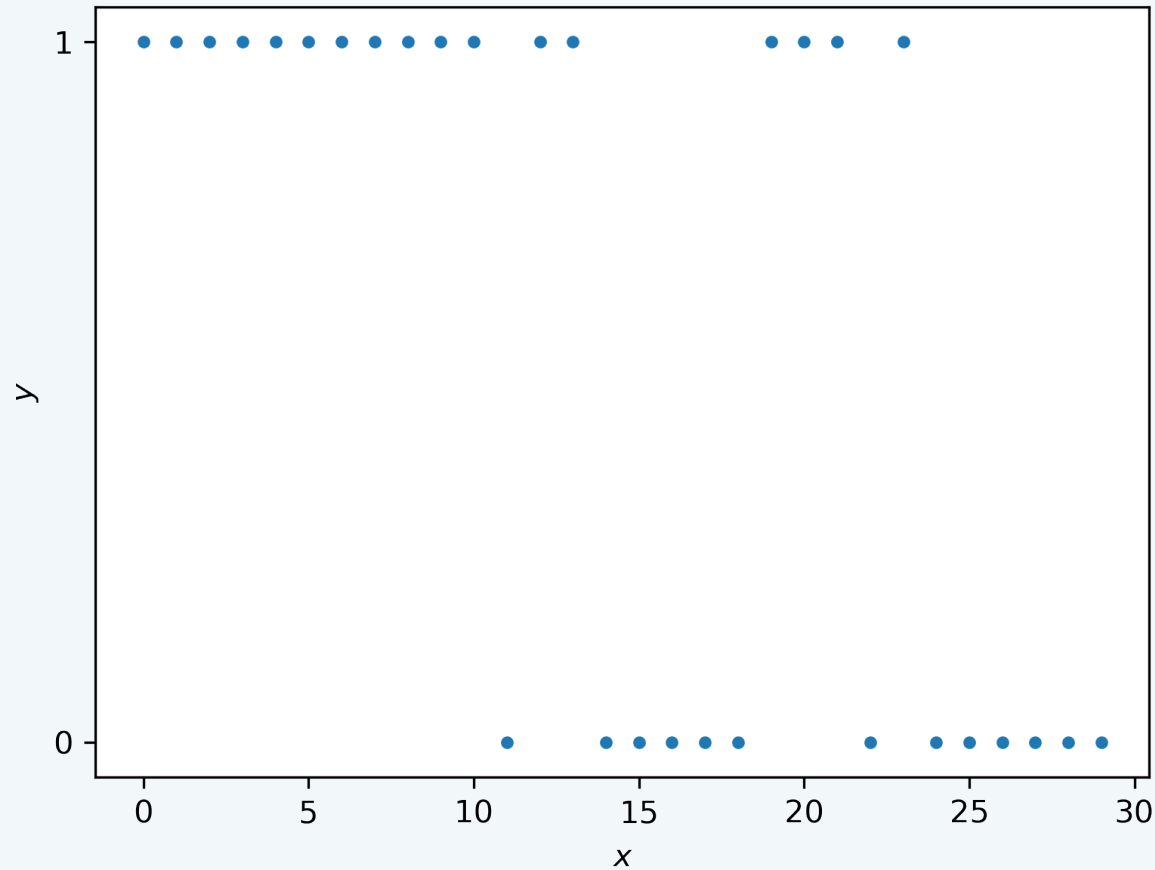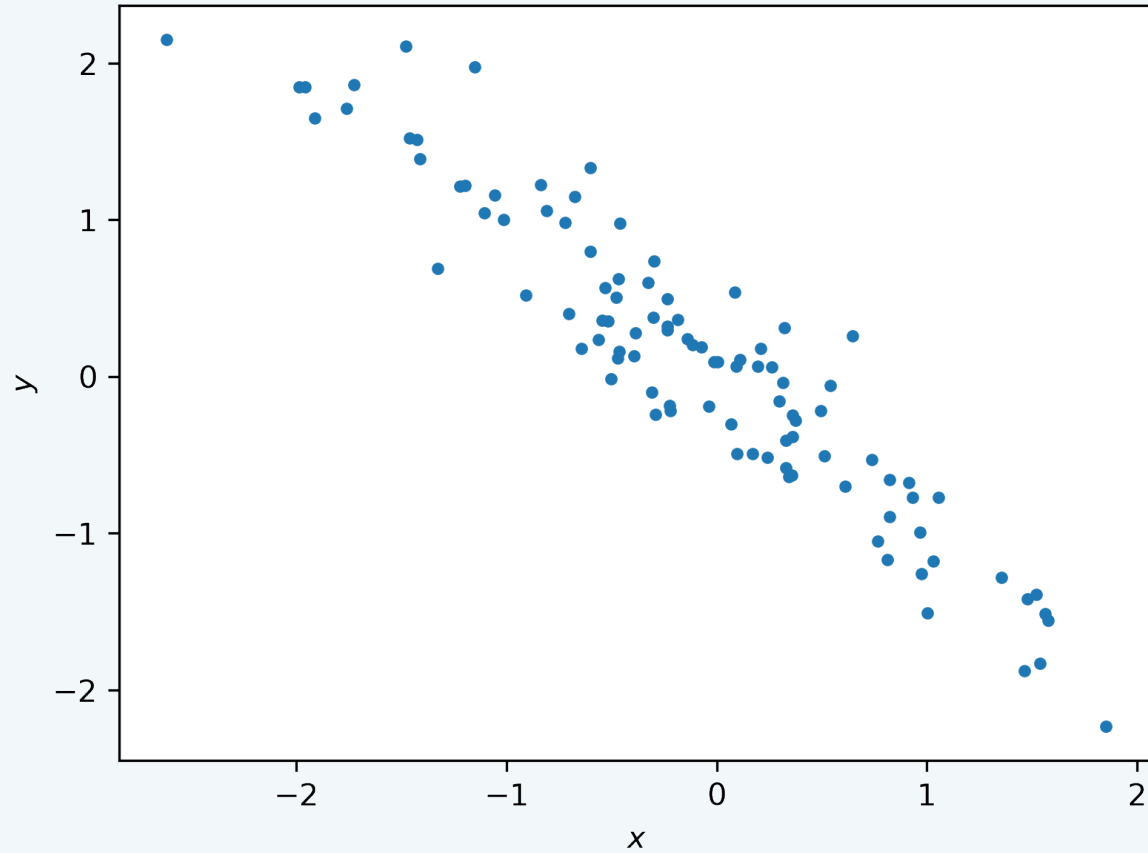- What can we do differently?

Example: Shooting Baskets

- Let's try shooting baskets from lots of different distances
- What effect would you expect? What do you see?
- Can we apply the "transformation" approach here?
- The two y values would map into two different y values
- What can we do differently?
- Let's think some more about regression and come back to this
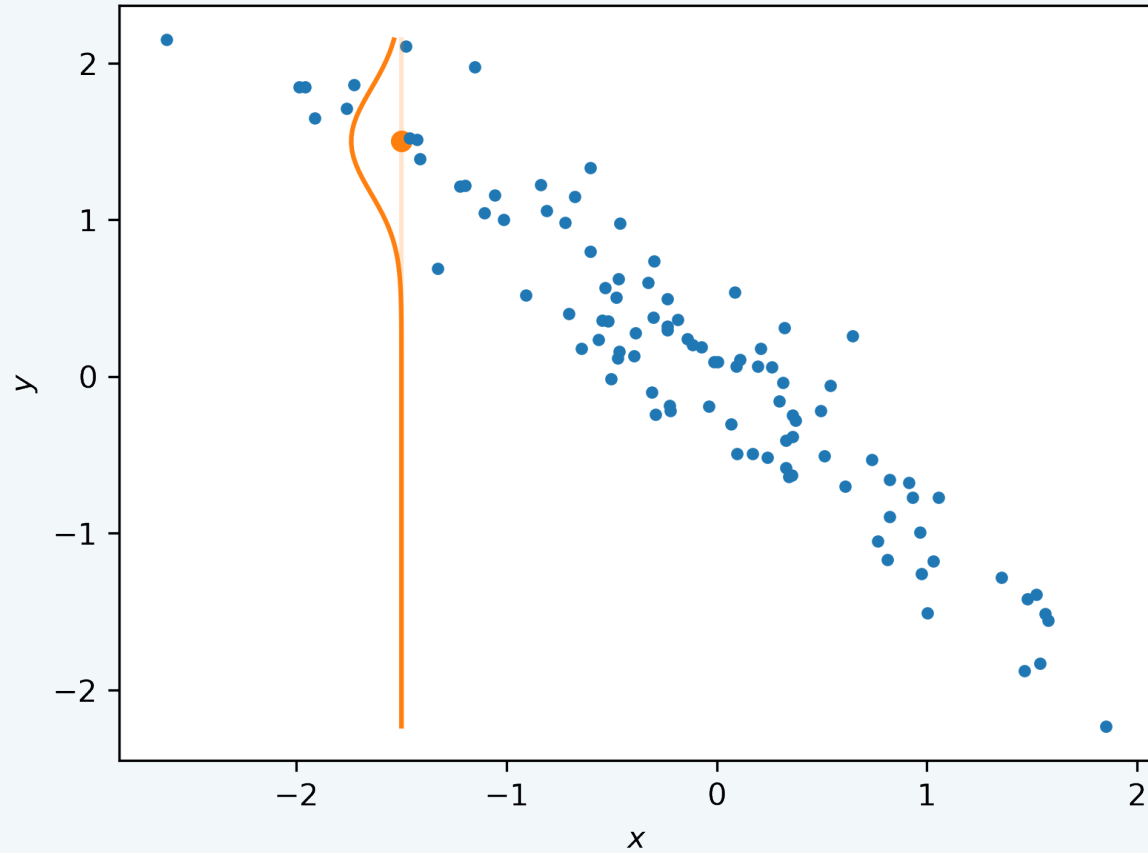
More Regression

- Classically, "regression" is about conditional expectation
- Do you see a conditional expectation here?
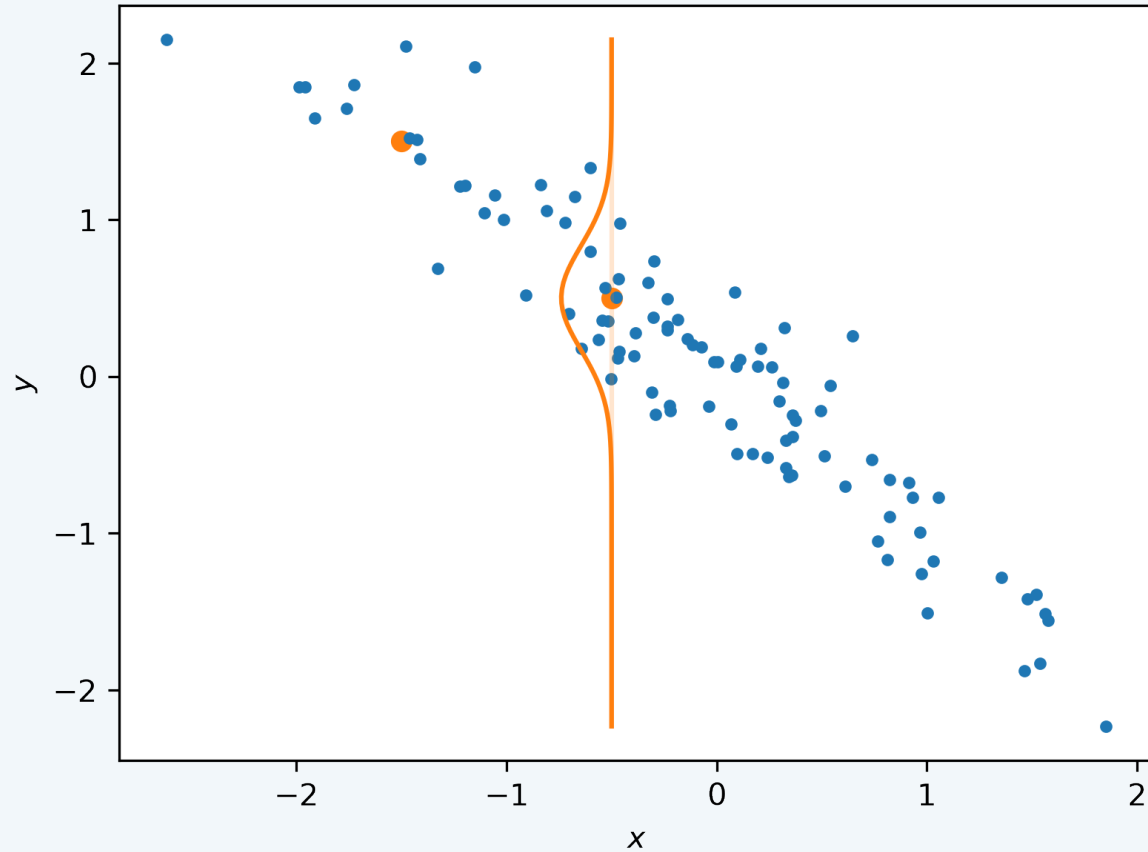
More Regression

- Classically, "regression" is about conditional expectation
- Do you see a conditional expectation here?
- For each x, we get a distribution over y

More Regression

- Classically, "regression" is about *conditional expectation*

- Do you see a conditional expectation here?

- For each *x*, we get a distribution over *y*

- For any *x*, our prediction is $\hat{y} = \mathbb{E}(y \mid x)$

More Regression

- Classically, "regression" is about *conditional expectation*

- Do you see a conditional expectation here?

- For each *x*, we get a distribution over *y*

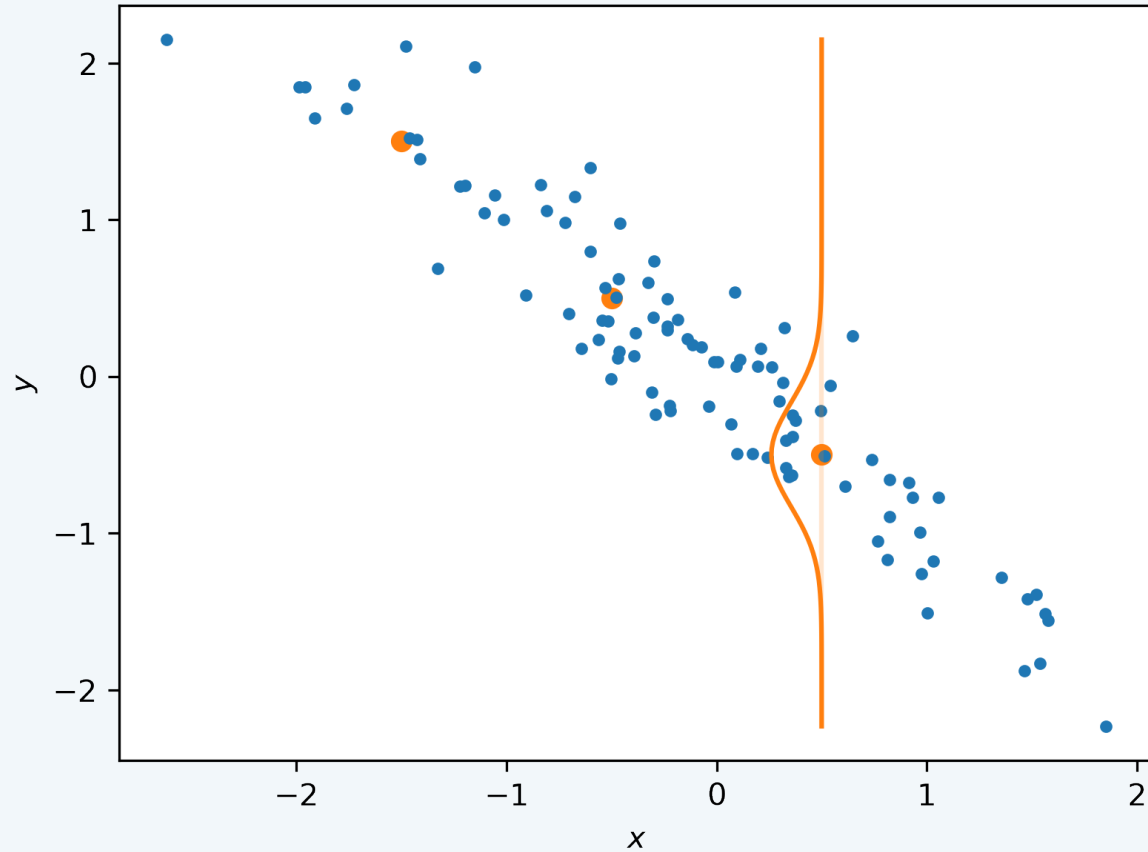- For any *x*, our prediction is $\hat{y} = \mathbb{E}(y \mid x)$

More Regression

- Classically, "regression" is about *conditional expectation*

- Do you see a conditional expectation here?

- For each *x*, we get a distribution over *y*

- For any *x*, our prediction is $\hat{y} = \mathbb{E}(y \mid x)$
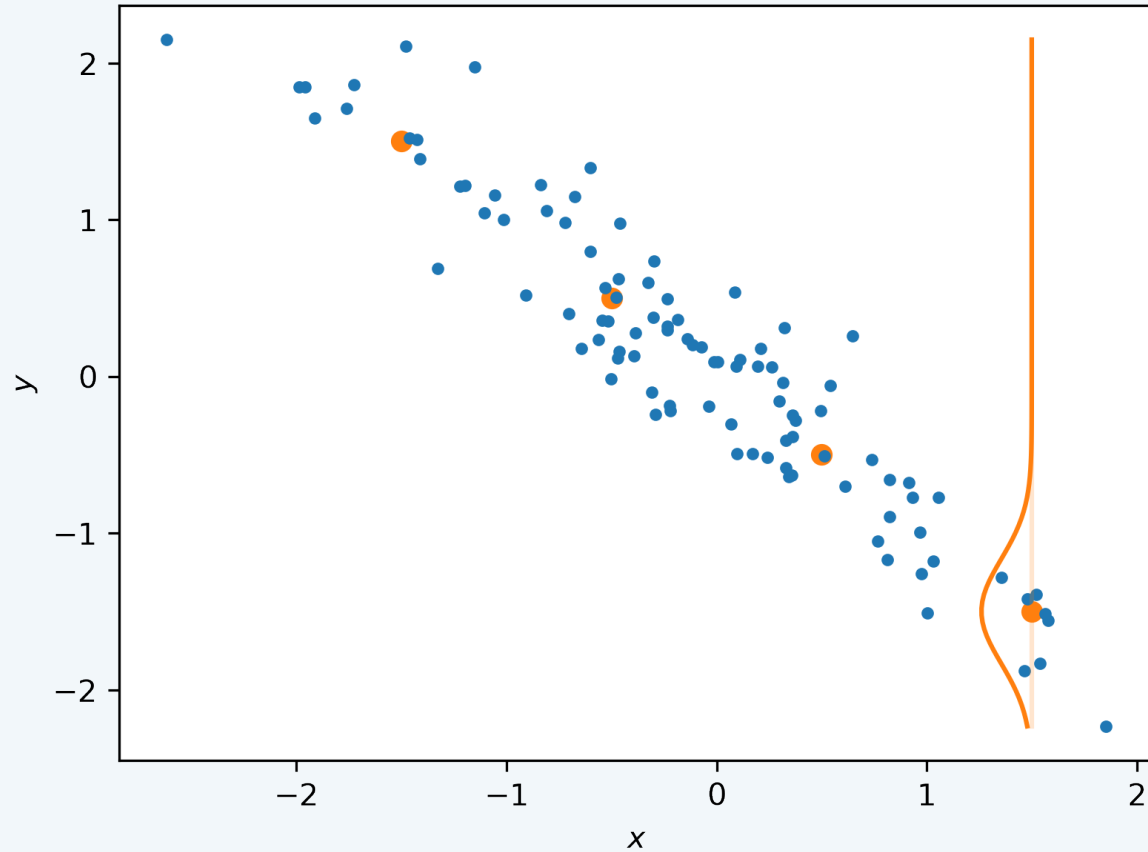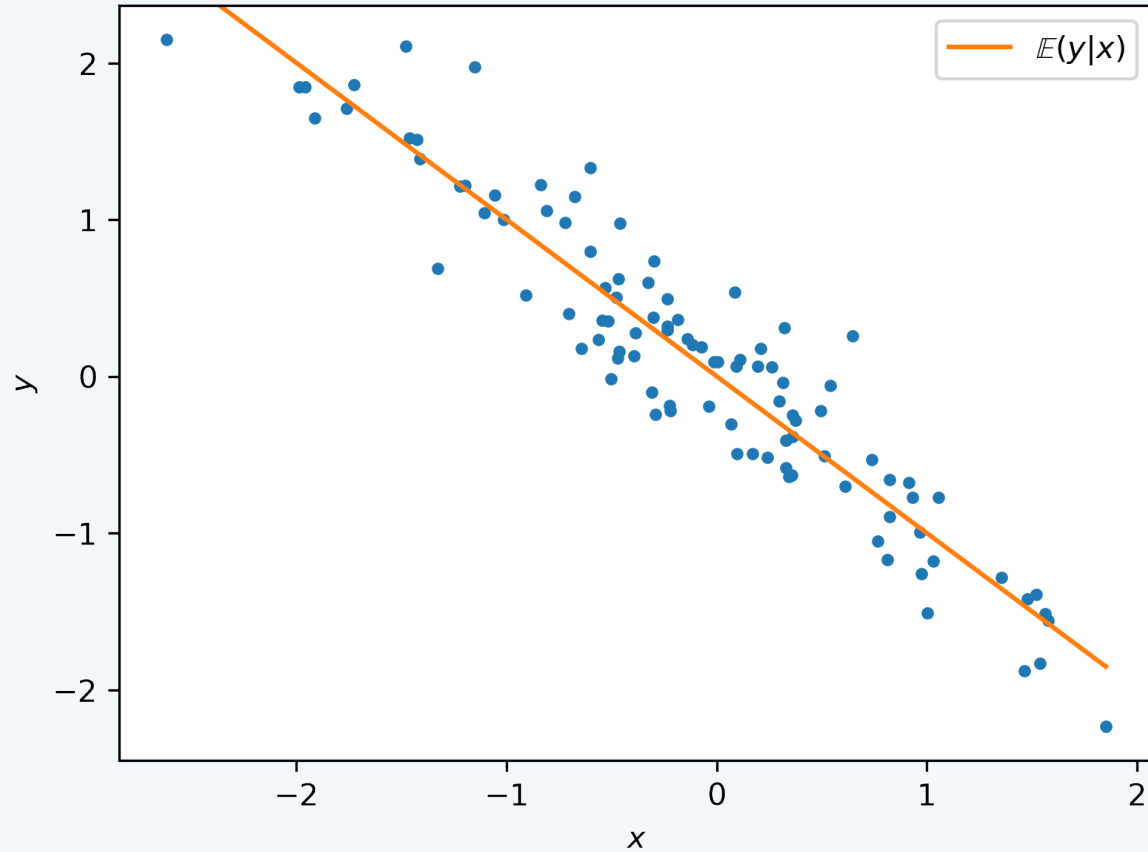
- Connecting these over all *x* gives us a curve
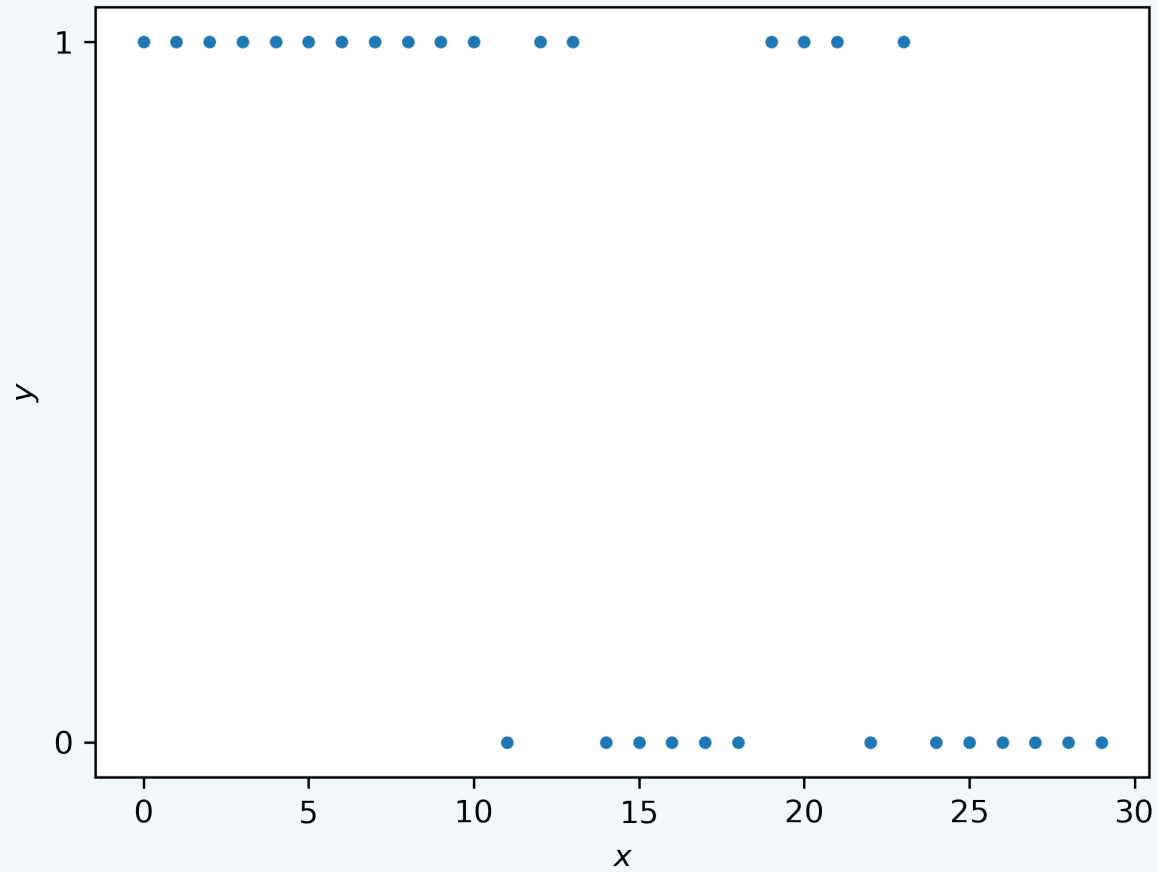
More Regression

- Classically, "regression" is about *conditional expectation*

- Do you see a conditional expectation here?

- For each *x*, we get a distribution over *y*

- For any *x*, our prediction is $\hat{y} = \mathbb{E}(y \mid x)$

- Connecting these over all *x* gives us a curve

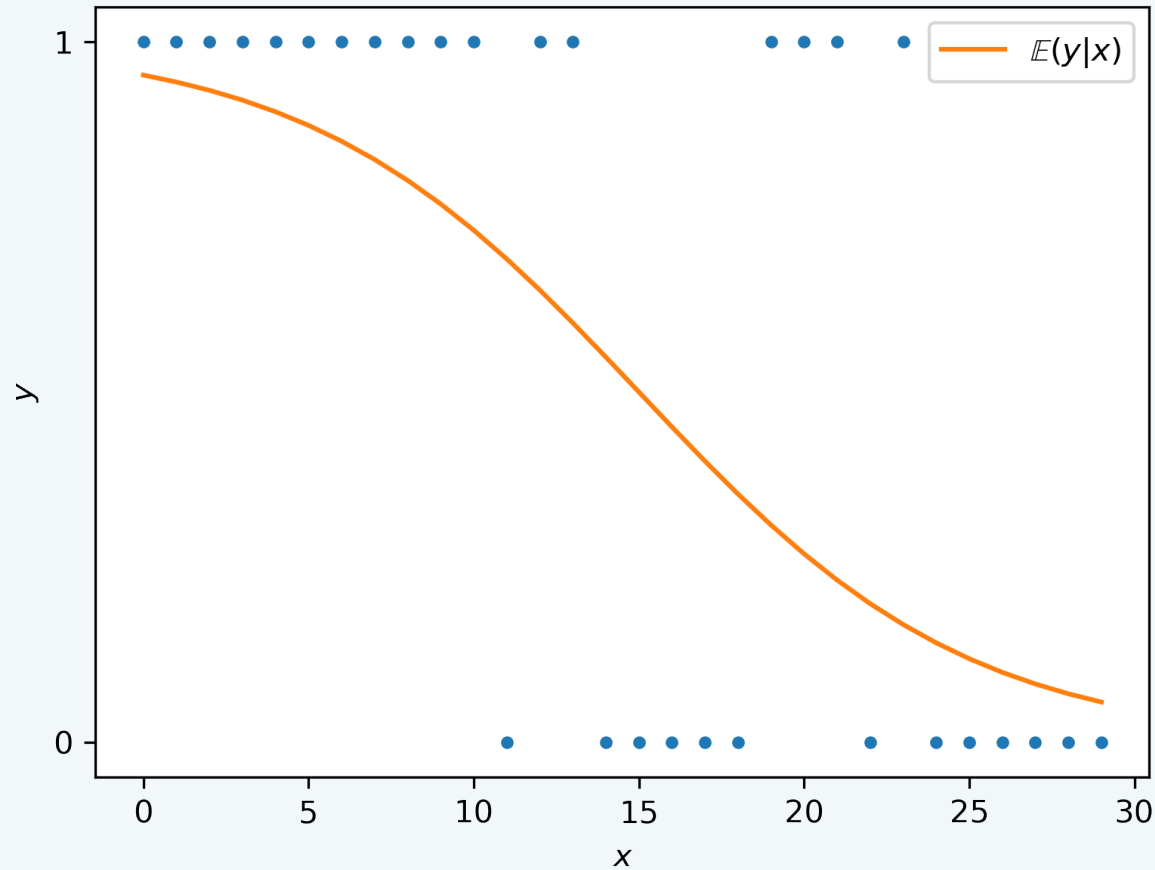- For linear regression, this "conditional expectation" curve is linear (in the parameters)

Back to Baskets

- For each x, y follows a Bernoulli ("biased coin flip") distribution

- What might conditional expectation look like here?

Back to Baskets

- For each *x*, *y* follows a Bernoulli ("biased coin flip") distribution

- What might conditional expectation look like here?

- In this case "conditional expectation" is also a conditional probability,

$$\mathbb{E}(y\,|\,x) = P(y = 1\,|\,x)$$

- Hey, that's what we need for classification!

- But... how do we get there from *x*?

Finding the Link
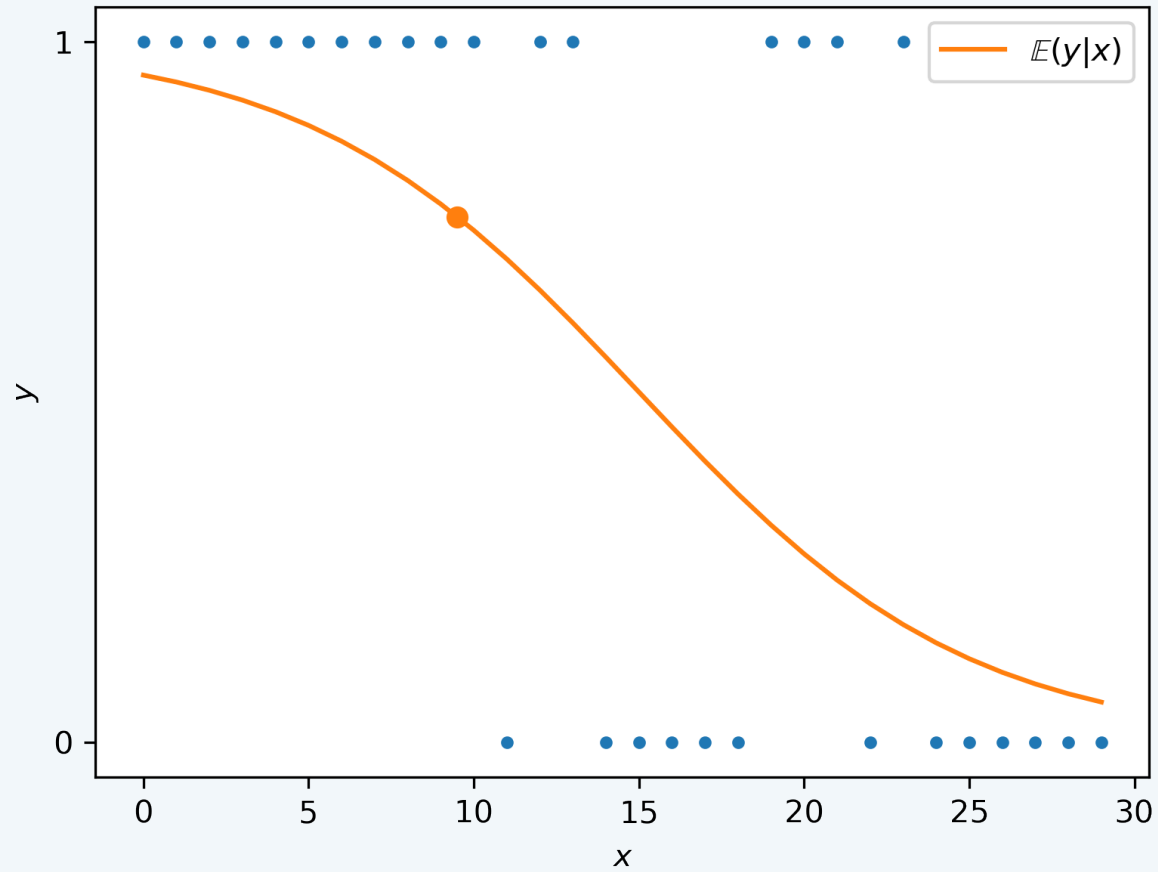
- For linear modeling, our model was

$$y = X\beta + \varepsilon$$

$$\varepsilon \sim \text{Normal}(0, \sigma)$$

- We used the data to find the best $\hat{\beta}$

- For a new "row vector" x, our prediction was $\mathbb{E}(y \mid x) = x\hat{\beta}$

- In that case the mean and linear predictor were the same

- For the baskets example, we'll need some transformation to "link" them

Back to Baskets

- Here's an x with $\mathbb{E}(y\,|\,x) = 3\big/4$
- Probabilities are always in $(0,1)$

Back to Baskets
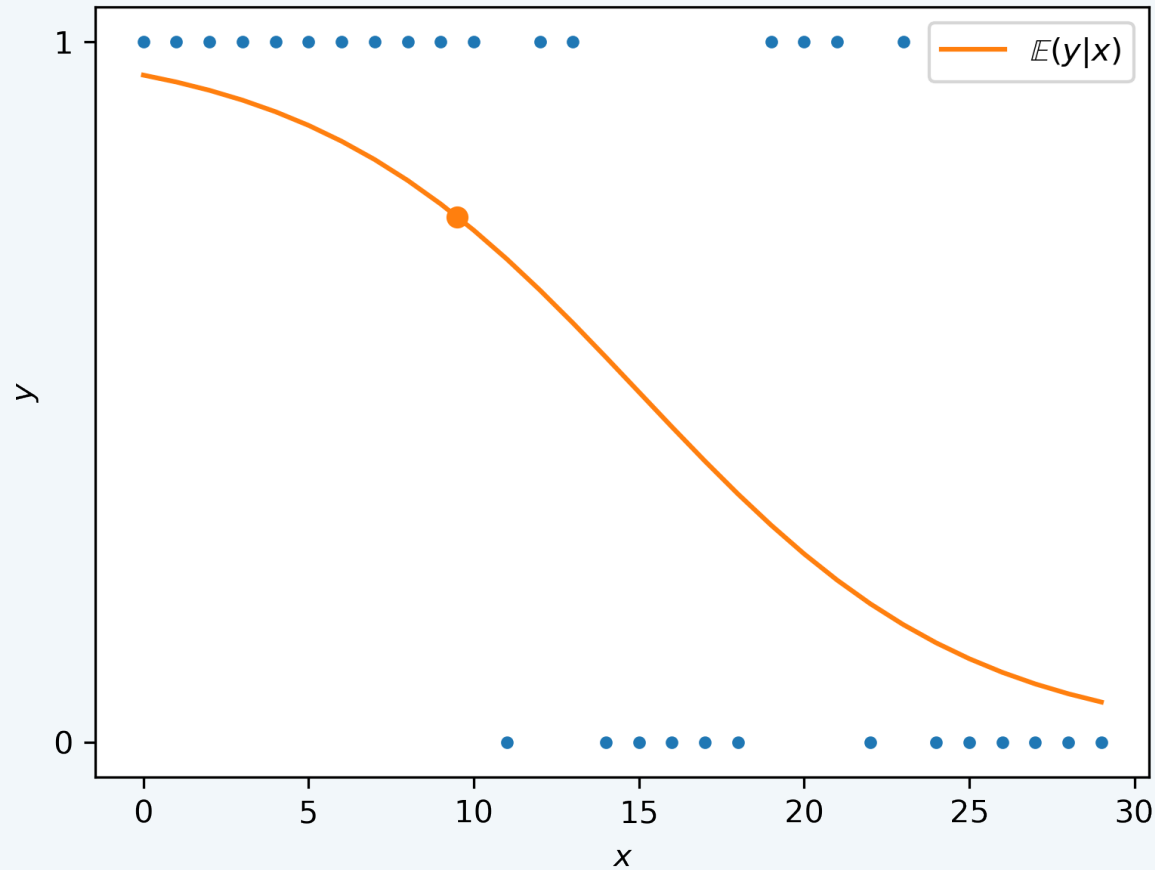
- Here's an *x* with $\mathbb{E}(y\,|\,x) = 3\big/4$

- Probabilities are always in (0,1)

- We can also work in terms of *odds*, which are 3:1 or 3/1

- If the mean is p, odds are $\dfrac{p}{1-p}$

- What values are possible for odds?

Back to Baskets

- Here's an *x* with $\mathbb{E}(y\,|\,x) = 3/4$

- Probabilities are always in (0,1)

- We can also work in terms of *odds*, which are 3:1 or 3/1

- If the mean is p, odds are $\dfrac{p}{1-p}$

- What values are possible for odds?

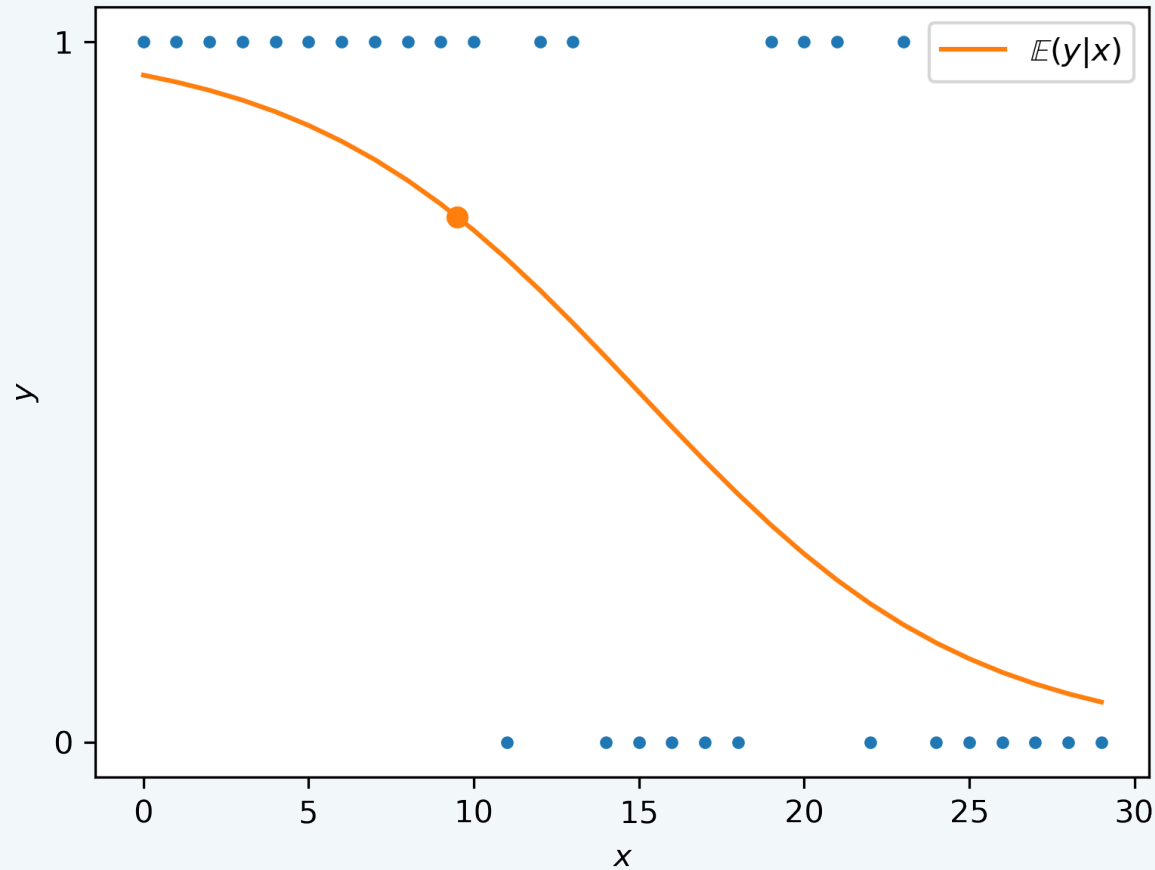- Odds can be any positive number!

Back to Baskets

- Here's an *x* with $\mathbb{E}(y\,|\,x) = 3\big/4$

- Probabilities are always in (0,1)

- We can also work in terms of *odds*, which are 3:1 or 3/1

- If the mean is p, odds are $\dfrac{p}{1-p}$
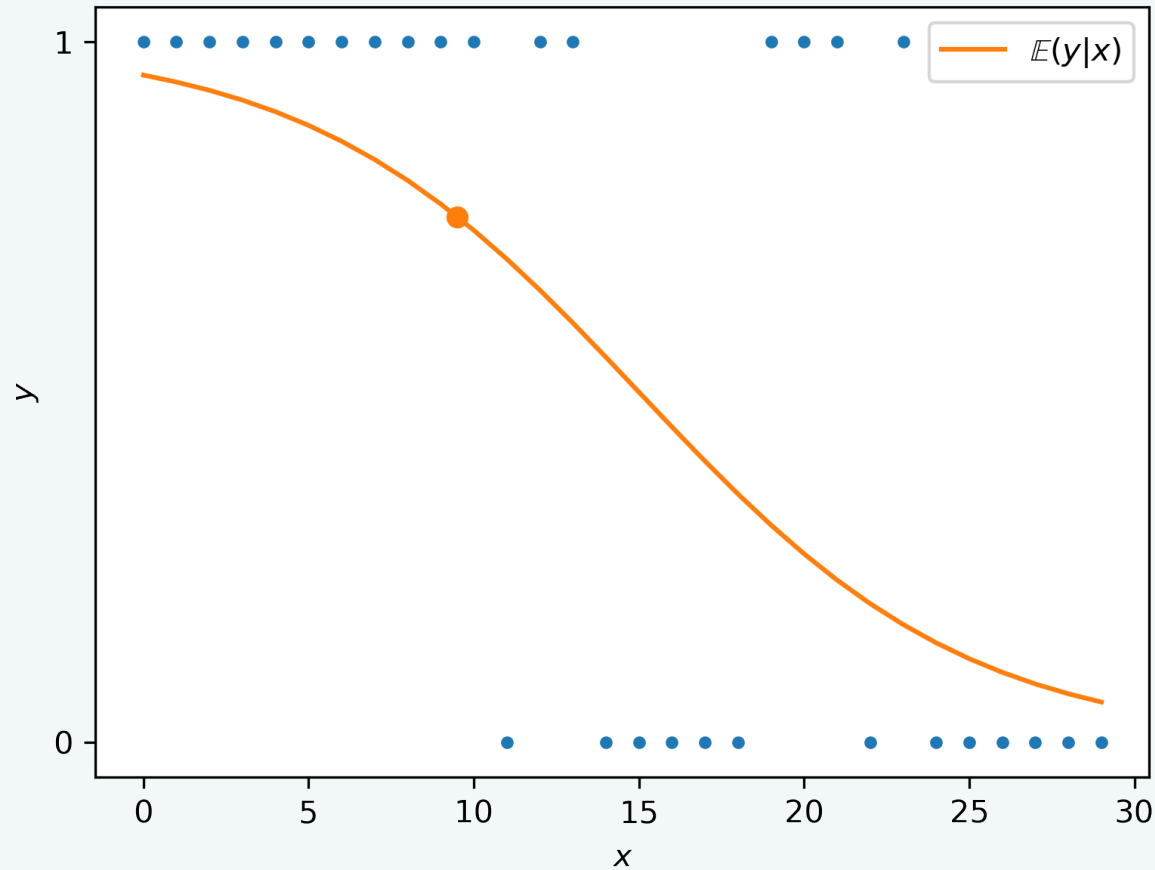
- What values are possible for odds?

- Odds can be any positive number!

- But our linear predictor can be any real number

- How can we get from "positives" to "reals"?

Back to Baskets

- Here's an *x* with $\mathbb{E}(y\,|\,x) = 3/4$

- Probabilities are always in (0,1)

- We can also work in terms of *odds*, which are 3:1 or 3/1

- If the mean is p, odds are $\dfrac{p}{1-p}$

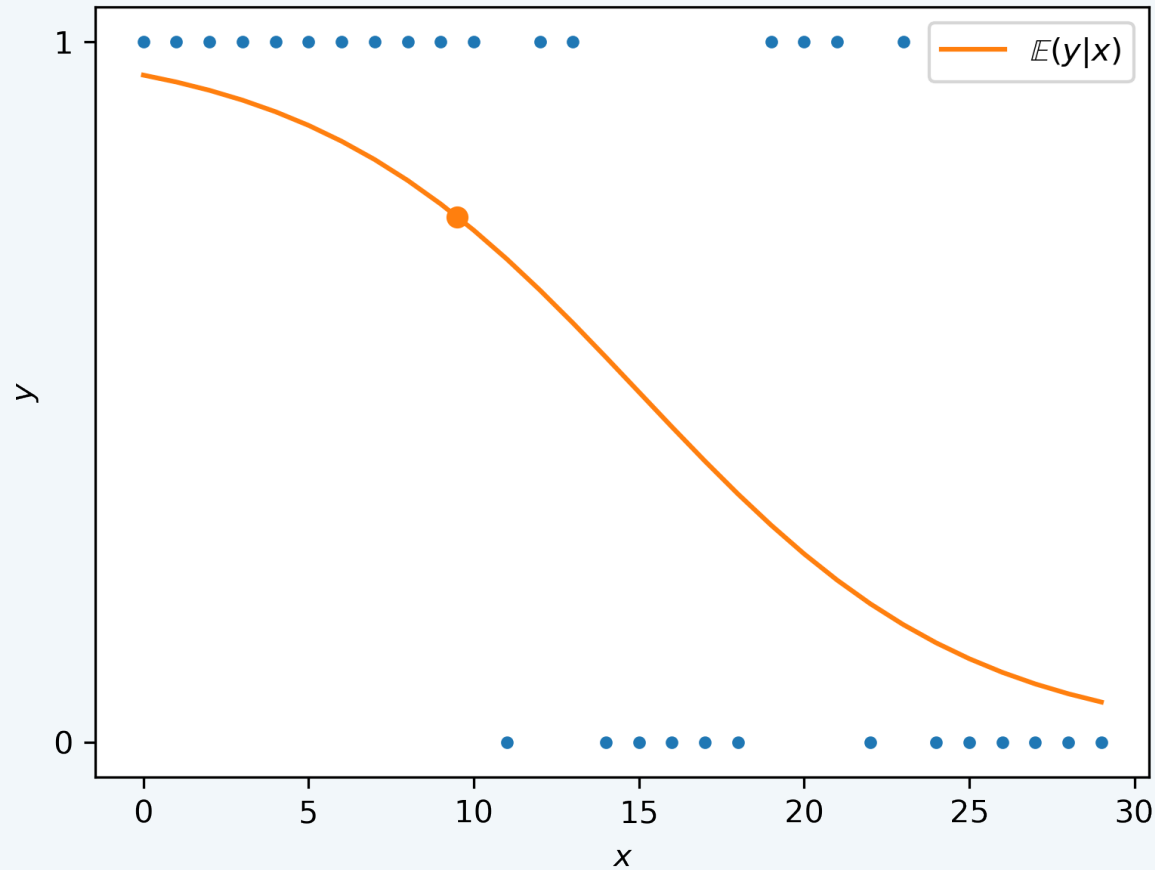- What values are possible for odds?

- Odds can be any positive number!

- But our linear predictor can be any real number

- How can we get from "positives" to "reals"?

- Use the log-odds!

$$p \mapsto \log \frac{p}{1-p}$$

The Logit

▶ The *logit* ("low jit") maps a probability to its log-odds

$$\text{logit } p = \log \frac{p}{1-p}$$

▶ For a normal linear model, the linear predictor $x\hat{\beta}$ gets us directly to $\mathbb{E}(y|x)$

▶ For logistic regression, it gives us a transformed version, $x\hat{\beta} = \text{logit } \mathbb{E}(y|x)$

▶ The logit "links" the linear predictor to the conditional expectation

▶ **For logistic regression, the logit is the *link function***

▶ (We'll see other link functions another time)
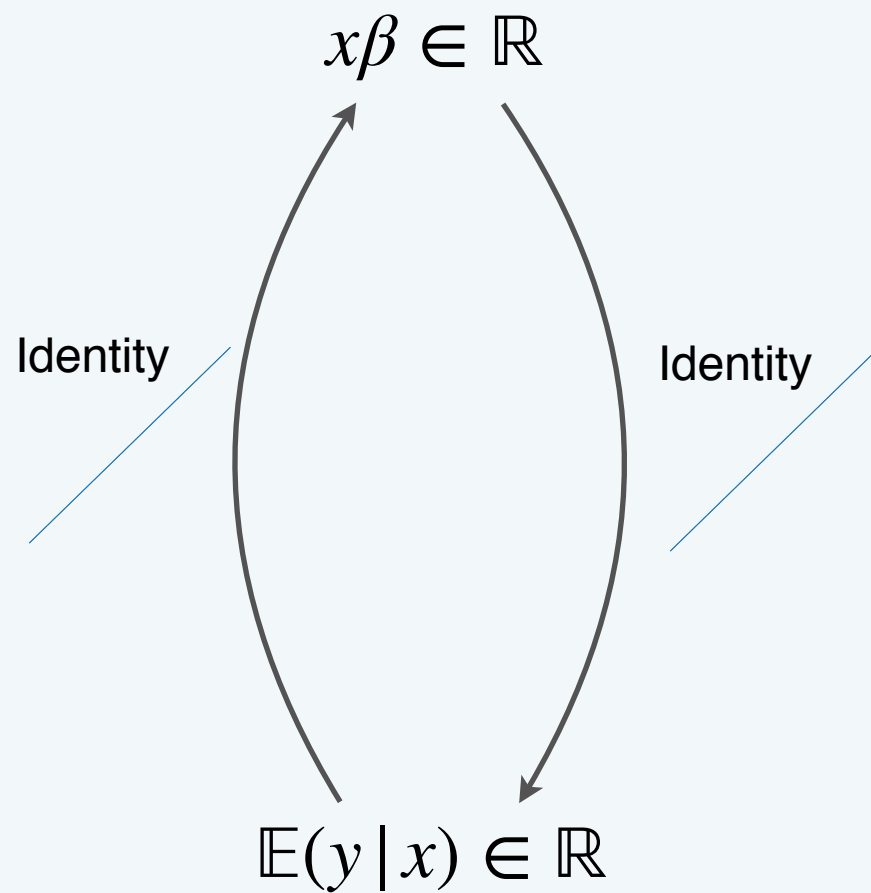
A Link and its Inverse

These mean the same thing:

$$\text{logit}\left(\mathbb{E}(y\,|\,x)\right) = x\beta$$

$$\mathbb{E}(y\,|\,x) = \text{logit}^{-1}(x\beta)$$
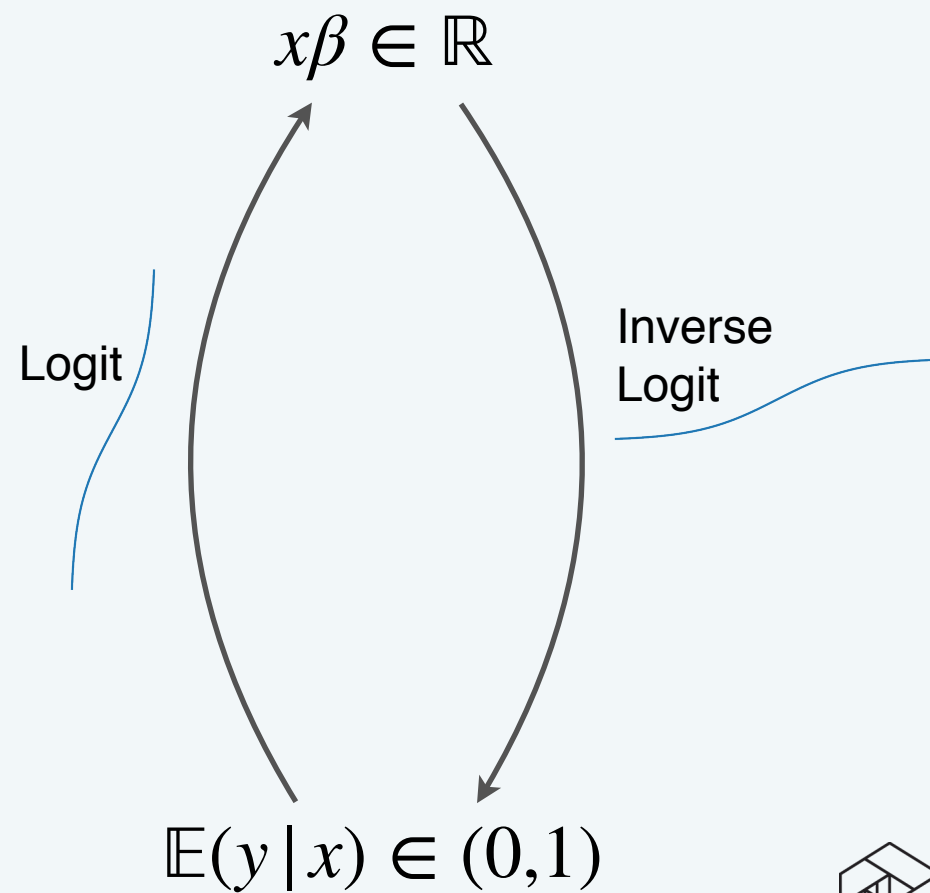
A Tale of Two Models

Linear Regression

$$x\beta \in \mathbb{R}$$

Identity

Identity

$$\mathbb{E}(y \mid x) \in \mathbb{R}$$

Logistic Regression

$$x\beta \in \mathbb{R}$$

Logit

Inverse Logit

$$\mathbb{E}(y \mid x) \in (0,1)$$

Sorry for the Synonyms

These are all the same thing:

▶ Inverse logit

▶ logistic function

▶ expit (in scipy)

▶ sigmoid (this means "S-shaped", but people usually mean this specifically)

$$x \mapsto \frac{1}{1 + e^{-x}}$$

Final Thoughts

▶ Loosely speaking, logistic regression is "linear regression for probabilities"

▶ Two differences:

   ▶ Distribution is binomial instead of normal

   ▶ Link function is logit instead of identity

▶ Many ideas from linear regression apply:

   ▶ Feature engineering

   ▶ Regularization (L_1, L_2, Elastic net)

   ▶ Diagnostics (*p*-values, log-likelihood, etc)