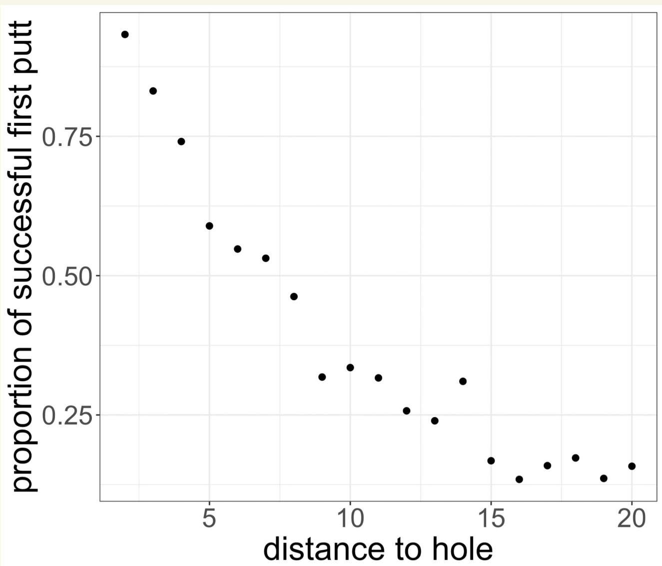
Logistic Regression

Q Estimate the probability that a putt is sunk as a function of distance to hole

Dataset of 5,988 putts from Mark Brodie include distance to hole and whether putt was sunk.



1/x

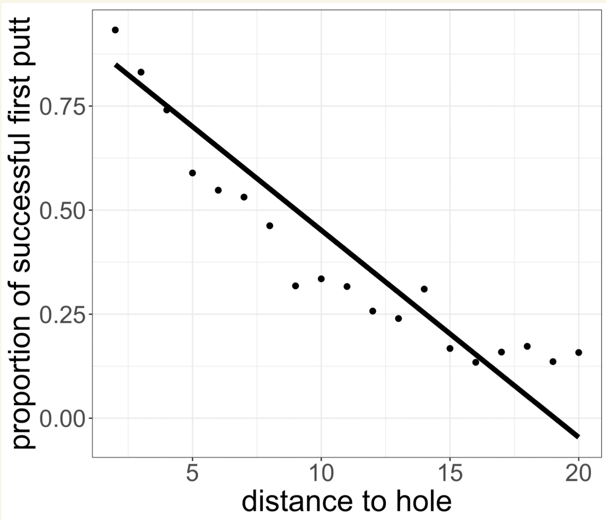
not linear

variables i = index of i^{th} putt

X_i = dist = dist to hole > 0

Y_i = $succ_i$ = 1 if putt sunk, 0 if not

Model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$



not as good as we can do

not linear

Model $Y_i = \beta_1 \frac{1}{X_i} + \varepsilon_i$

$\mathbb{E} Y_i = \beta_0 + \cancel{\beta_1 \frac{1}{X_i}} = p_i = 0$ and $X_i = \infty$

$$Y_i = \begin{cases} 1 & \text{if make putt} \\ 0 & \text{if not} \end{cases} = \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{w.p. } 1-p_i \end{cases}$$

$$\mathbb{E} Y_i = \sum_{y \in \{0,1\}} y \, \mathbb{P}(Y_i = y)$$

$$= 1 \cdot \mathbb{P}(Y_i = 1) + 0 \cdot \mathbb{P}(Y_i = 0)$$

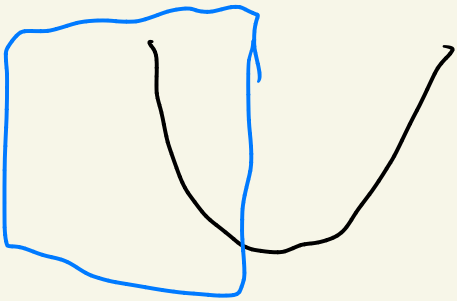
$$= \mathbb{P}(Y_i = 1) = p_i$$

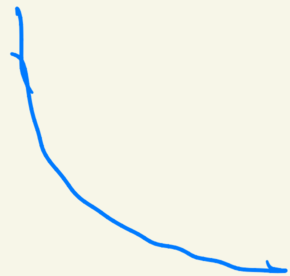
$$\left\{ \begin{array}{l} \text{if } X_i \approx 0 \text{ then } \frac{1}{X_i} \approx \infty \\ \quad \text{so } \quad Y_i \approx \infty \\ \quad \text{but } \quad Y_i = 1 \text{ or } 0 \end{array} \right.$$

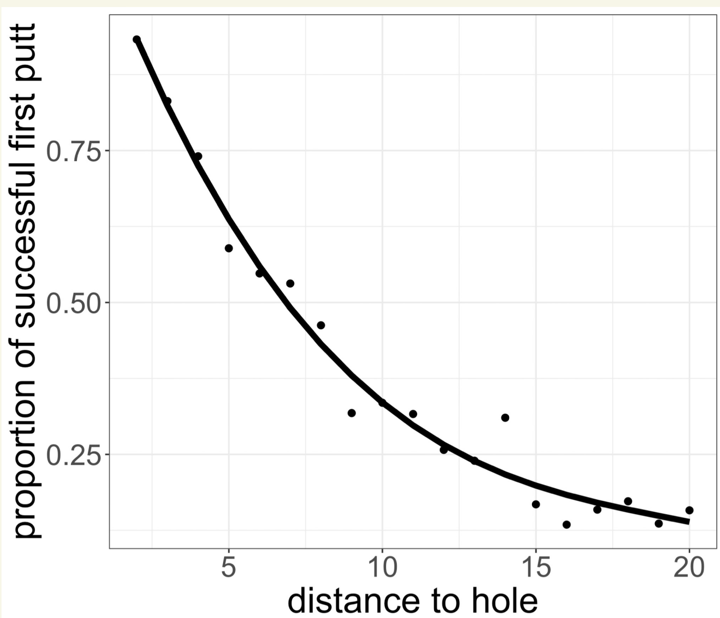
- diff. model
- want to make $\hat{Y}_i \in [0,1]$

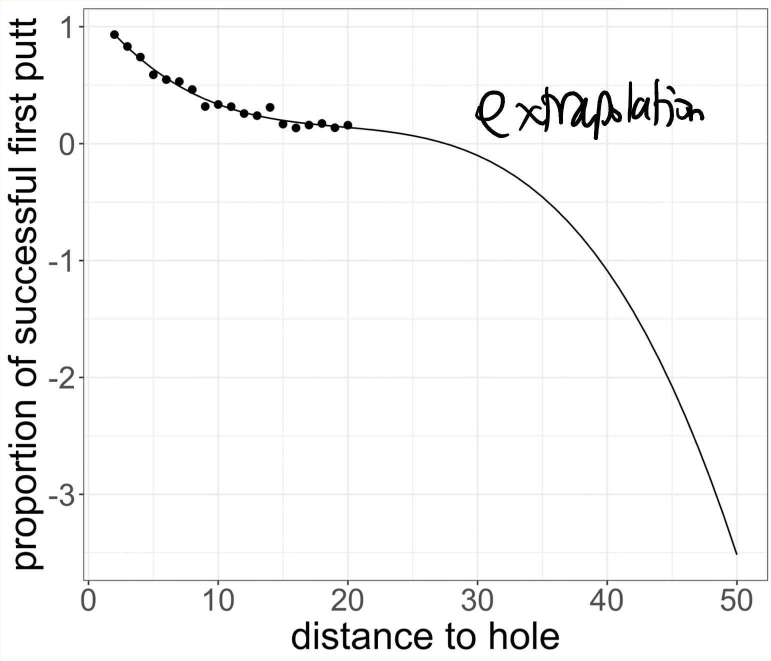
Model $\quad Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i$

Model $\quad Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i$









* Find something that converges to 0 as $X \to \infty$

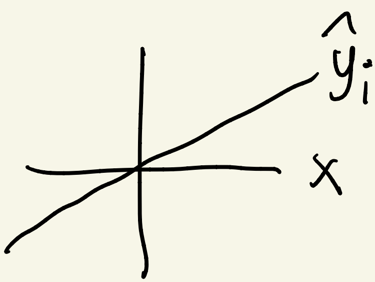
* Force our prediction $\hat{Y}_i$ to be in $[0,1]$.
  Force our model to lie in $[0,1]$, for $x \in [0, \infty)$.

Problem the probability of an event must lie in $[0,1]$, ordinary linear regression model does not ~~assumf~~ guarantee ~~for~~ this.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Linear Regression $\quad y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

How can we force our predictions to be in $[0, 1]$.



$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

the predicted values $\hat{y}_i$ can be any number in $(-\infty, \infty)$ depending on the value of X_i

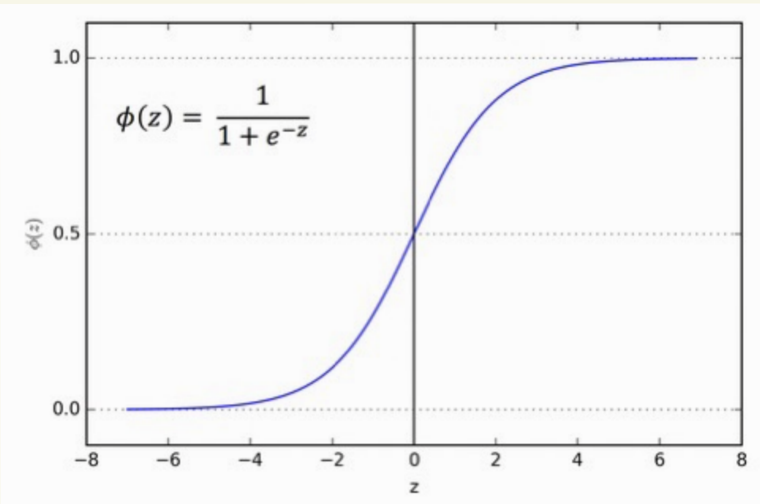
Squishification Function

↳ takes a number in $(-\infty, \infty)$ and squishes it into $[0,1]$.

$$\text{Logistic}(z) = \frac{1}{1+e^{-z}} = \text{Sigmoid}(z)$$

$$= \sigma(z)$$

$$\approx \phi(z)$$



$z=\infty$, $\quad \frac{1}{1+e^{-z}} = \frac{1}{1+0} = 1$

$z=-\infty$, $\quad \frac{1}{1+e^{-z}} = \frac{1}{1+\infty} = 0$

$z=0$, $\quad \frac{1}{1+e^{-z}} = \frac{1}{1+1} = 1/2$

before $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

sqnish $\hat{y}_i = \text{Logistic}(\hat{\beta}_0 + \hat{\beta}_1 x_i)$

$$= \frac{1}{1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 x_i)}}$$

$$\hat{y}_i = \text{Logistic}(x_i^T \beta)$$

$$x_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{pmatrix} = \begin{pmatrix} x_i \\ x_i^2 \\ x_i^3 \\ \vdots \\ x_i^n \end{pmatrix}$$

$$\hat{y}_i = \text{Logistic}(\hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2)$$

$$\hat{p}_i = \mathbb{P}(y_i = 1) = \hat{y}_i = \frac{1}{1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 x_i)}}$$

Model

$$p_i = P(y_i = 1) = \frac{1}{1 + e^{-(x_i^T \beta)}}$$

$$y_i \sim \text{Bernoulli}(p_i) = \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{w.p. } 1 - p_i \end{cases}$$

$y_i = 1, 1, 0, 0, 0, 1, 0, 1, 0$

x_i^T Row vector

$x_i^T = (1, x_i) \rightarrow x_i^T \beta = \beta_0 + \beta_1 x_i$

$x_i^T = (1, x_i, x_i^2, x_i^3) \rightarrow x_i^T \beta = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3$

$\hat{\beta}_0, \hat{\beta}_1$

How to estimate the coefficients?

LR: $\underset{\beta}{\text{argmin}}\ RSS(\beta) = \sum_{i=1}^{n} (y_i - \underbrace{x_i^T \beta}_{\hat{y}_i})^2$

Logistic Reg: $y_i = 0$ or 1

Minimize <u>log loss</u>, a.k.a <u>cross entropy loss</u>

$$\boxed{\begin{aligned} L(\beta) &= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log P_i + (1-y_i)\log(1-P_i)\right] \\ \text{where} \quad P_i &= \mathbb{P}(y_i = 1 \mid x_i, \beta) = \frac{1}{1+e^{-x_i^T\beta}} \end{aligned}}$$

$x_i = dist_i$ $\qquad$ $x_i = (1, dist_i, dist_i^2, dist_i^3)$

$$L_i(\beta) = -y_i \log P_i - (1-y_i)\log(1-P_i).$$

When $y_i = 1$, $L_i(\beta) = -\log P_i$

when $P_i \approx 1$, $L_i \approx 0$

when $P_i \approx 0$, $L_i \approx \infty$

when $y_i = 0$, $L_i(\beta) = -\log(1 - p_i)$

when $p_i \approx 1$, $L_i \approx \infty$

when $p_i \approx 0$, $L_i \approx 0$

* Let's minimize the logloss to find our coeffs:

$$\underset{\beta}{\operatorname{argmin}} \; L(\beta) \qquad \beta = (\beta_0, \beta_1, \ldots, \beta_K)$$

$$0 = \nabla_\beta L(\beta)$$

$$= \nabla_\beta \left\{ -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log p_i + (1 - y_i) \log(1 - p_i) \right] \right\}$$

$$= -\frac{1}{n} \sum_{i=1}^{n} \left\{ y_i \left[\nabla_\beta \log p_i \right] + (1 - y_i) \left[\nabla_\beta \log(1 - p_i) \right] \right\}$$

$$p_i = \frac{1}{1 + e^{-(x_i^T \beta)}} = \sigma(x_i^T \beta)$$

$$\nabla_\beta \log P_i = \left(\frac{\partial}{\partial \beta_0} \log P_i, \ldots, \frac{\partial}{\partial \beta_k} \log P_i \right)$$

$$\frac{\partial}{\partial \beta_j} \log P_i$$

$$= \frac{\partial}{\partial \beta_j} \log \left(\frac{1}{1+e^{-(x_i^T \beta)}} \right)$$

$$= \frac{\partial}{\partial \beta_j} \left[-\log \left(1 + e^{-(x_i^T \beta)} \right) \right]$$

$$= \frac{-\frac{\partial}{\partial \beta_j} \left(1 + e^{-(x_i^T \beta)} \right)}{1+e^{-(x_i^T \beta)}}$$

$$-(x_{0i}\beta_0 + x_{1i}\beta_1 + \ldots + x_{ki}\beta_k)$$

$$= \frac{-e^{-x_i^T \beta} \frac{\partial}{\partial \beta_j} \left(-x_i^T \beta \right)}{1+e^{-(x_i^T \beta)}}$$

$$\frac{\partial}{\partial \beta_j} \log P_i = \left(\frac{e^{-x_i^T \beta}}{1 + e^{-x_i^T \beta}} \right) \cdot x_{ji}$$

$\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_K)$

$$\frac{\partial}{\partial \beta_0} \log P_i = \left(\frac{e^{-x_i^T \beta}}{1 + e^{-x_i^T \beta}} \right) \cdot x_{0i}$$

$$\frac{\partial}{\partial \beta_1} \log P_i = \left(\frac{e^{-x_i^T \beta}}{1 + e^{-x_i^T \beta}} \right) \cdot x_{1i}$$

$$\frac{\partial}{\partial \beta_2} \log P_i = \left(\frac{e^{-x_i^T \beta}}{1 + e^{-x_i^T \beta}} \right) \cdot x_{2i}$$

$x_i^T \beta = (1, x_i, x_i^2, x_i^3)^T \cdot (\beta_0, \beta_1, \beta_2, \beta_3)$
$= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3$

$x_i^T = (x_{0i}, x_{1i}, \ldots, x_{ki})$

$$\nabla_\beta \log P_i = \left(\frac{\partial}{\partial \beta_0} \log P_i, \ldots, \frac{\partial}{\partial \beta_K} \log P_i \right)$$

$$= \left(\frac{e^{-x_i^T \beta}}{1 + e^{-x_i^T \beta}} \right) \cdot (x_{0i}, x_{1i}, \ldots, x_{ki})$$

$$= \frac{e^{-x_i^T\beta}}{1+e^{-x_i^T\beta}} \cdot x = (1-p_i)x_i$$

$$p_i = \frac{1}{1+e^{-(x_i^T\beta)}}, \quad 1-p_i = \cancel{\frac{1+e^{-x_i^T\beta}}{1+e^{-x_i^T\beta}}} - \frac{\cancel{1}}{1+e^{-x_i^T\beta}}$$

$$= \frac{e^{-x_i^T\beta}}{1+e^{-x_i^T\beta}}$$

$$\begin{cases} \nabla_\beta \log p_i = (1-p_i)x_i \\ \nabla_\beta \log(1-p_i) = -p_i x_i \end{cases}$$

$$\nabla_\beta L(\beta) =$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left\{y_i\left[\nabla_\beta \log p_i\right] + (1-y_i)\left[\nabla_\beta \log(1-p_i)\right]\right\}$$

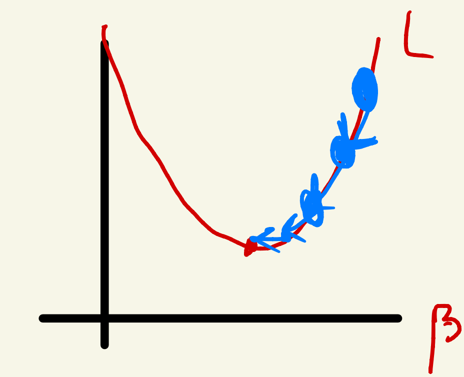
$$= -\frac{1}{n}\sum_{i=1}^{n}\left\{y_i(1-p_i)x_i - (1-y_i)p_i x_i\right\}$$

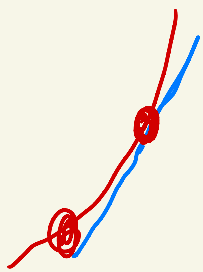
$$= -\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)x_i$$

$$= -\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \sigma(x_i^T\beta)\right)\cdot x_i = 0$$

No known closed form solution for β.

Gradient Descent





descend down the gradient until convergence:

$$\beta^{(t+1)} \leftarrow \beta^{(t)} + K \cdot \nabla_\beta L(\beta^{(t)})$$

$$\boxed{\beta^{(t+1)} \leftarrow \beta^{(t)} + k \cdot \sum_{i=1}^{n} (y_i - p_i) x_i}$$

Stop when $\| \beta^{(t)} - \beta^{(t+1)} \| < \delta$.

.000100000001

ELO

One iteration of gradient descent in logistic regression is equivalent to one ELO update

everyone has their own ELO rating/~~score~~/strength score parameter β.

player A plays against player B.

β_A vs. β_B

$$p_{AB} = \mathbb{P}(A \text{ beats } B) = \mathbb{P}(y_{AB} = 1)$$

$$:= \frac{1}{1 + e^{-(\beta_A - \beta_B)}} = \sigma(\beta_A - \beta_B)$$

$$\boxed{\begin{aligned} \beta_A &\leftarrow \beta_A + K(y_{AB} - p_{AB}) \\ \beta_B &\leftarrow \beta_B + K(y_{BA} - p_{BA}) \end{aligned}}$$

$y_{AB} = 1$ if A beats B else 0

$p_{AB} = \mathbb{P}(A \text{ beats } B)$

- If A beats B $(y_{AB} = 1)(y_{BA} = 0)$

  If $p_{AB} \approx 1$ $(p_{BA} \approx 0) \longrightarrow \beta_A$ barely increases

  If $p_{AB} \approx 0$ $(p_{BA} \approx 1) \longrightarrow \beta_A$ increases by K

K = K factor = learning rate

$$P_{AB} = \frac{1}{1+e^{-(\beta_A - \beta_B)}}$$

- golf putting example

estimate putt prob. as a function of dist to hole

set up logistic regression model

$$p_i = \mathbb{P}(y_i = 1) = \frac{1}{1+e^{-(x_i^T \beta)}}$$

$$x_i = (1, \text{dist}_i)$$

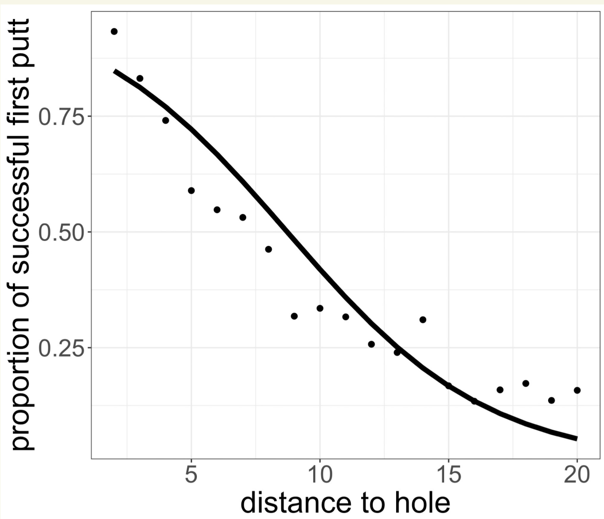
Ran gradient descent to minimize $L(\beta)$ ← logloss

from this, estimate $\beta \longrightarrow \hat{\beta}$

then, go back and use $\hat{p} = \frac{1}{1+e^{-x^T\hat{\beta}}}$

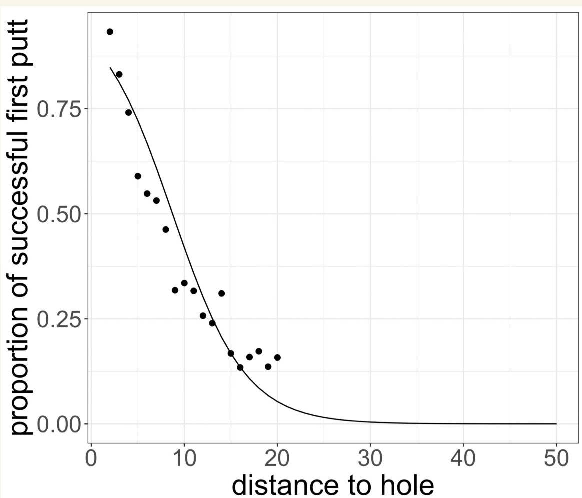
Let's visualize $\hat{p}$ from logistic regression:

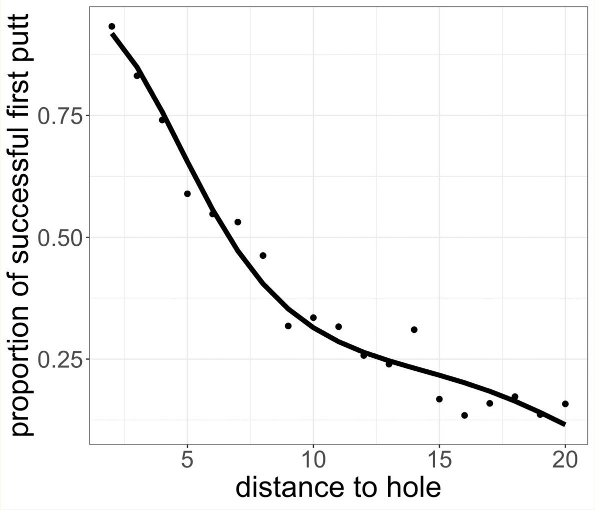
- decide to use logistic regression ($y_i \in \{0,1\}$)
- acquire (x_i, y_i)
- run 1 line of R code



Looks decent

dist_i





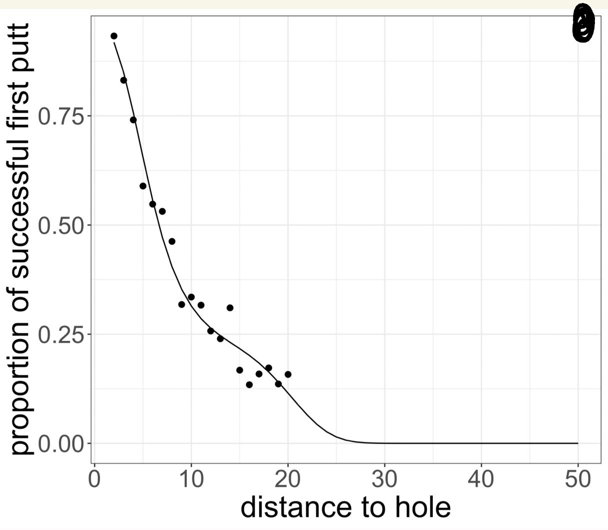
$$x_i^T = (1, x_i, x_i^2, x_i^3)$$

d, r

d

d, r^2

d, r^3



Confounder:

elevation

hole placement

Bradley Terry Power Scores

Logistic Regression Power Scores

Schedule matrix X from yesterday

game i, Home team $H(i)$, Away team $A(i)$,

$$X_{ij} = \begin{cases} 1 & \text{if } j = \text{intercept column} \\ 1 & \text{if } j = H(i) \\ -1 & \text{if } j = A(i) \\ 0 & \text{else} \end{cases}$$

Outcomes win/loss y

$$y_i = \begin{cases} 1 & \text{if } H(i) \text{ wins} \\ 0 & \text{if } H(i) \text{ loses} \end{cases}$$

Model

$$p_i = \mathbb{P}(y_i = 1) = \frac{1}{1 + e^{-x_i^T \beta}}$$

$$= \frac{1}{1 + e^{-(\beta_{H(i)} - \beta_{A(i)} + \beta_0)}}$$

$$y_i \sim \text{Bernoulli}(p_i)$$