Logistic Regression Analysis

Dr. Daniel B. Rowe
Professor of Computational Statistics
Department of Mathematical and Statistical Sciences
Marquette University

Outline

Logistic Regression Background

Logistic Regression Model

Logistic Regression Estimation

Logistic Regression Example

Logistic Regression Discussion

Homework

Logistic Regression Background

Often the probability p of an event E depends upon an independent variable x, such as the probability p of getting an A on the final depends on the number of hours that you study x.

i.e. as x increases so does p.

So p is a function of x, $p(x)$.

Hours (x)	A (y)
6	0
8	0
10	0
12	0
14	0
16	1
18	0
20	0
22	0
24	0
26	1
28	0
30	0
32	1
34	1
36	1
38	1
40	1

Logistic Regression Background

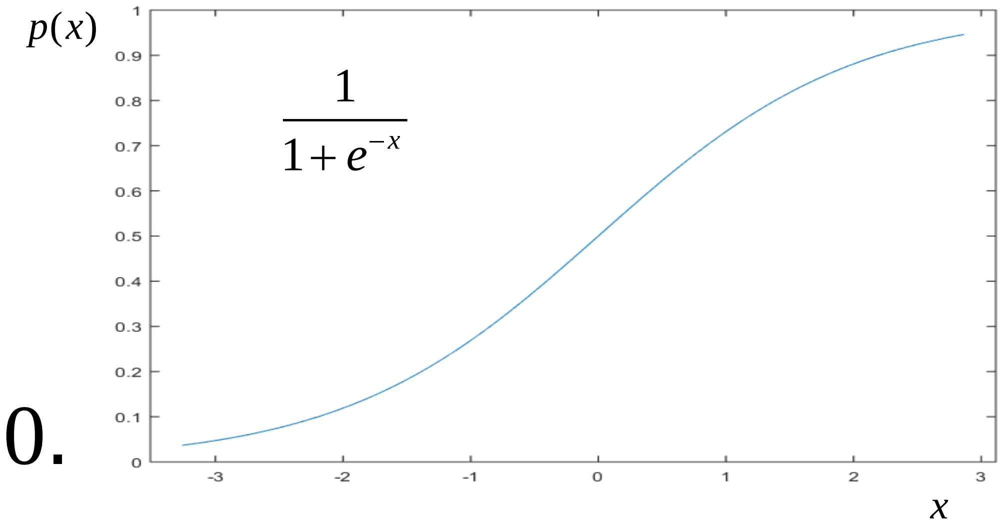
If you study x=10 hours then your probability $p(x)$ of getting an A might be $p(10)=0.25$,
but if you study x=30 hours then your probability $p(x)$ of getting an A might be $p(30)=0.75$
i.e. as x increases so does p.
So p is a function of x, $p(x)$!

Hours (x)	A (y)
6	0
8	0
10	0
12	0
14	0
16	1
18	0
20	0
22	0
24	0
26	1
28	0
30	0
32	1
34	1
36	1
38	1
40	1

Logistic Regression Model

This dependency of a probability $p(x)$, $0 \leq p(x) \leq 1$, on an independent variable $x, -\infty < x < \infty$, is generally described through the logistic mapping function

$$p = p(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} \quad .$$



If the event E occurs, then we say y=1 and if not y=0. $P(y=1)=p$ and $P(y=0)=1\text{-}p$. …

This is a Bernoulli trial whose probability of success depends on x.

Verhulst, 1838; Ostwald, 1883; .., Fisher,1935,….

Logistic Regression Model

The PMF for a Bernoulli trial is

$$P(Y=y)=p^{y}(1-p)^{1-y}\ ,\ \ y=\{0,1\},\ \ 0\leq p\leq 1.$$

If we have a collection of n independent Bernoulli trials where the probability of success is different for each,

$$P(Y_1=y_1)\cdots P(Y_n=y_n)=\prod_{i=1}^{n}p_i^{y_i}(1-p_i)^{1-y_i}\ ,\ \ y_i=\{0,1\},\ \ 0\leq p_i\leq 1$$

is the likelihood (joint PMF).

Logistic Regression Model

The Bernoulli likelihood,

$$P(Y_1 = y_1)\cdots P(Y_n = y_n) = \prod_{i=1}^{n} [p(x_i)]^{y_i} [1 - p(x_i)]^{1-y_i}$$

written with the p_i's as a function of the corresponding x_i's

$$p(x_i) = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_i}}$$

becomes

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} [p(x_i)]^{y_i} [1 - p(x_i)]^{1-y_i}$$

.

Logistic Regression Estimation

So how do we estimate β_0 and β_1?

The probability for all n is called the likelihood function

$$L(\beta_0,\beta_1)=\prod_{i=1}^{n}[p(x_i)]^{y_i}[1-p(x_i)]^{1-y_i} \quad \text{where} \quad p(x_i)=\frac{1}{1+e^{-\beta_0-\beta_1 x_i}} \quad .$$

Our goal is to find the values of β_0 and β_1 that maximize the likelihood.

We can equivalently maximize natural log of the likelihood

$$LL(\beta_0,\beta_1)=\ln[L(\beta_0,\beta_1)] \; .$$

Logistic Regression Estimation

The likelihood function

$$L(\beta_0,\beta_1)=\prod_{i=1}^{n}[p(x_i)]^{y_i}[1-p(x_i)]^{1-y_i} \quad \text{where} \quad p(x_i)=\frac{1}{1+e^{-\beta_0-\beta_1 x_i}}$$

has log likelihood function

$$\begin{aligned} LL(\beta_0,\beta_1) &= \sum_{i=1}^{n}\left[y_i\ln[p(x_i)]+(1-y_i)\ln[1-p(x_i)]\right] \\ &= \sum_{i=1}^{n}y_i\ln[p(x_i)]+\sum_{i=1}^{n}(1-y_i)\ln[1-p(x_i)] \\ &= \sum_{i=1}^{n}\ln[1-p(x_i)]+\sum_{i=1}^{n}y_i\ln[p(x_i)/(1-p(x_i))] \\ &= -\sum_{i=1}^{n}\ln[1+e^{\beta_0+\beta_1 x_i}]+\sum_{i=1}^{n}y_i[\beta_0+\beta_1 x_i] \end{aligned}$$

.

Logistic Regression Estimation

$$p(x_i) = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_i}}$$

Differentiating

$$LL(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 x_i) - \sum_{i=1}^{n} \ln(1 + e^{\beta_0 + \beta_1 x_i})$$

with respect to β_0 and β_1 then setting equal to zero yields

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}$$

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} \frac{x_i e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}$$

we can’t uniquely solve for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Logistic Regression Estimation

$$p(x_i) = \frac{1}{1 + e^{-\beta_0 - \beta_1 x_i}}$$

So we need to numerically maximize the Log Likelihood

$$LL(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 x_i) - \sum_{i=1}^{n} \ln(1 + e^{\beta_0 + \beta_1 x_i})$$

with respect to β_0 and β_1 using a method such as

Grid Search, *Gradient Descent*, *Newton's Method*, …

to get $\hat{\beta}_0$ and $\hat{\beta}_1$.

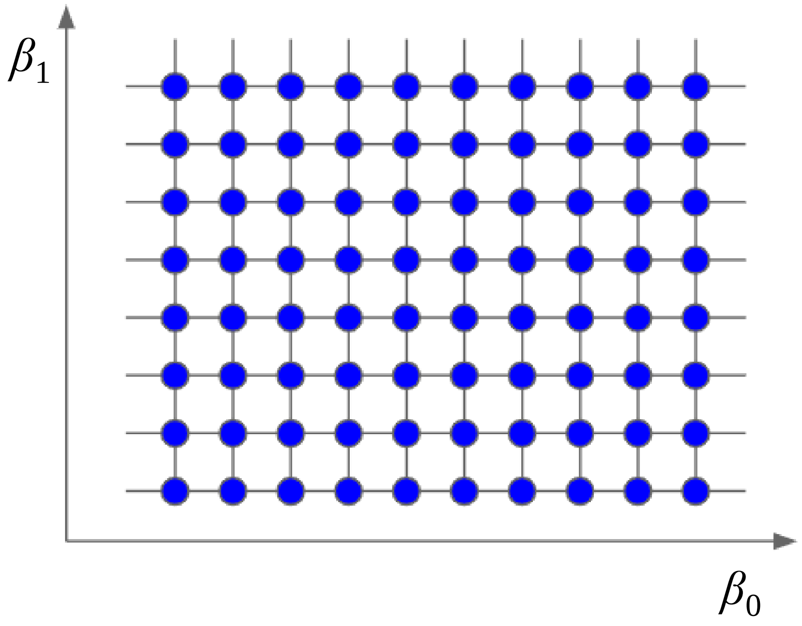
Logistic Regression Estimation

For a *Grid Search*, evaluate an array of combinations of β_0 and β_1 for

$$LL(\beta_0,\beta_1)=\sum_{i=1}^{n} y_i(\beta_0+\beta_1 x_i)-\sum_{i=1}^{n}\ln(1+e^{\beta_0+\beta_1 x_i})\ ,$$

β_0 between β_{0min} and β_{0max} in increments of say $.01$ and
β_1 between β_{1min} and β_{1max} in increments of say $.01$

the β_0 and β_1 that make LL the largest are $\hat{\beta}_0$ and $\hat{\beta}_1$.



Logistic Regression Estimation

For *Gradient Descent*, take derivatives with respect to β_0 and β_1,

$$LL(\beta_0,\beta_1)=\sum_{i=1}^{n} y_i(\beta_0+\beta_1 x_i)-\sum_{i=1}^{n}\ln(1+e^{\beta_0+\beta_1 x_i})$$

$$\frac{\partial LL}{\partial \beta_0}=\sum_{i=1}^{n} y_i-\sum_{i=1}^{n}\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$

$$\frac{\partial LL}{\partial \beta_1}=\sum_{i=1}^{n} x_i y_i-\sum_{i=1}^{n}\frac{x_i e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$$

move in β_0 and β_1 directions that increase LL until $\hat{\beta}_0$ and $\hat{\beta}_1$.

Logistic Regression Example

Student hours studied (x) to for A on final (y).

We relate hours studied x to binary outcome of earn A on final exam y, $0 \leq x$, $y = \{0,1\}$, $0 \leq p(x) \leq 1$.

Data simulated with true $\beta_0 = -5.0$, $\beta_1 = 0.2$,

success probabilities $p_i = \dfrac{1}{1 + e^{-\beta_0 - \beta_1 x_i}}$, $y_i \sim \mathrm{Bin}(1, p_i)$.

Estimate β_0 and β_1 by *Grid Search & Gradient Descent*.

Hours (x)	A (y)
6	0
8	0
10	0
12	0
14	0
16	1
18	0
20	0
22	0
24	0
26	1
28	0
30	0
32	1
34	1
36	1
38	1
40	1

Logistic Regression Example

Find the β_0 and β_1 combination that maximize the LL

$$LL(\beta_0,\beta_1)=\sum_{i=1}^{n} y_i(\beta_0+\beta_1 x_i)-\sum_{i=1}^{n}\ln(1+e^{\beta_0+\beta_1 x_i})$$

Set up a grid search. Try all combinations of β_0 between -10 and 10 in increments of .01 and β_1 between -5 and 5 in increments of .01.

Show code.

Hours (x)	A (y)
6	0
8	0
10	0
12	0
14	0
16	1
18	0
20	0
22	0
24	0
26	1
28	0
30	0
32	1
34	1
36	1
38	1
40	1

Logistic Regression Example

```
% Logistic Regression Simulation
rng('default')
% set true coefficients
beta0tru=-5; beta1tru= 0.2;
% select independent x values
dx=2; x=(6:dx:40)'; n=length(x);
% log odds of event
L=beta0tru+beta1tru*x;
% proportion of successes
p=exp(L)./(1+exp(L));
% generate dichotomous values
y=binornd(ones(n,1),p);
% Grid search values
beta0min=-10; beta0max=0;
beta1min= -5; beta1max=5;
beta0del=.01; beta1del=.01;
beta0=(beta0min:beta0del:beta0max)’;
beta1=(beta1min:beta1del:beta1max)';
```

```
m0=size(beta0,1);
m1=size(beta1,1);
% try all beta0, beta1 values
maxLL=-10^6; LL=ones(m0,m1);
beta0hat=0; beta1hat=0;
for i=1:m0
 for j=1:m1
 LL(i,j)=sum(y.*(beta0(i,1)...
         +beta1(j,1)*x))...
  -sum(log(1+exp(beta0(i,1)...
         +beta1(j,1)*x)));
  if (LL(i,j)>maxLL)
   beta0hat=beta0(i,1);
   beta1hat=beta1(j,1);
   maxLL=LL(i,j);
  end
 end
end
```

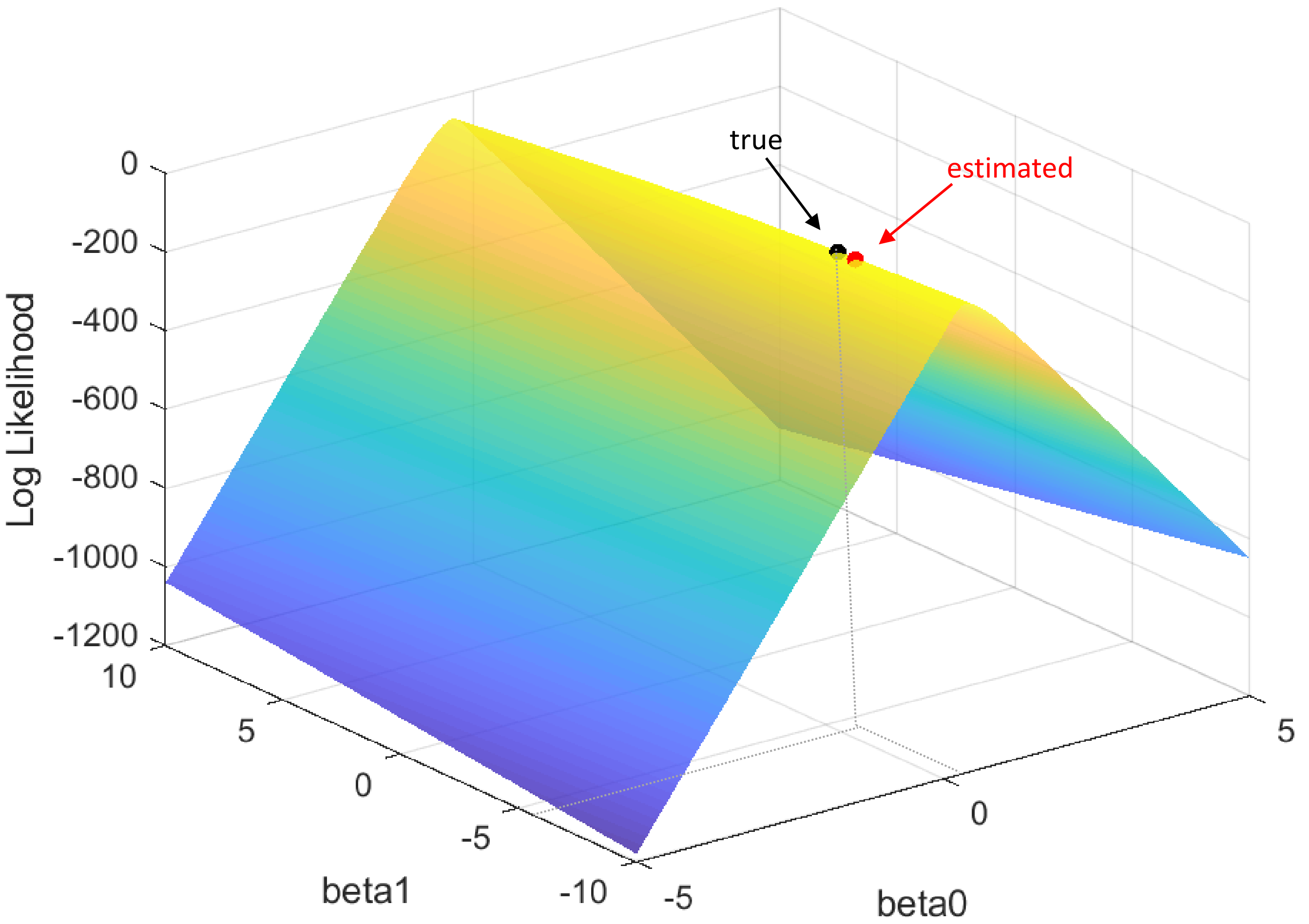
Logistic Regression Example

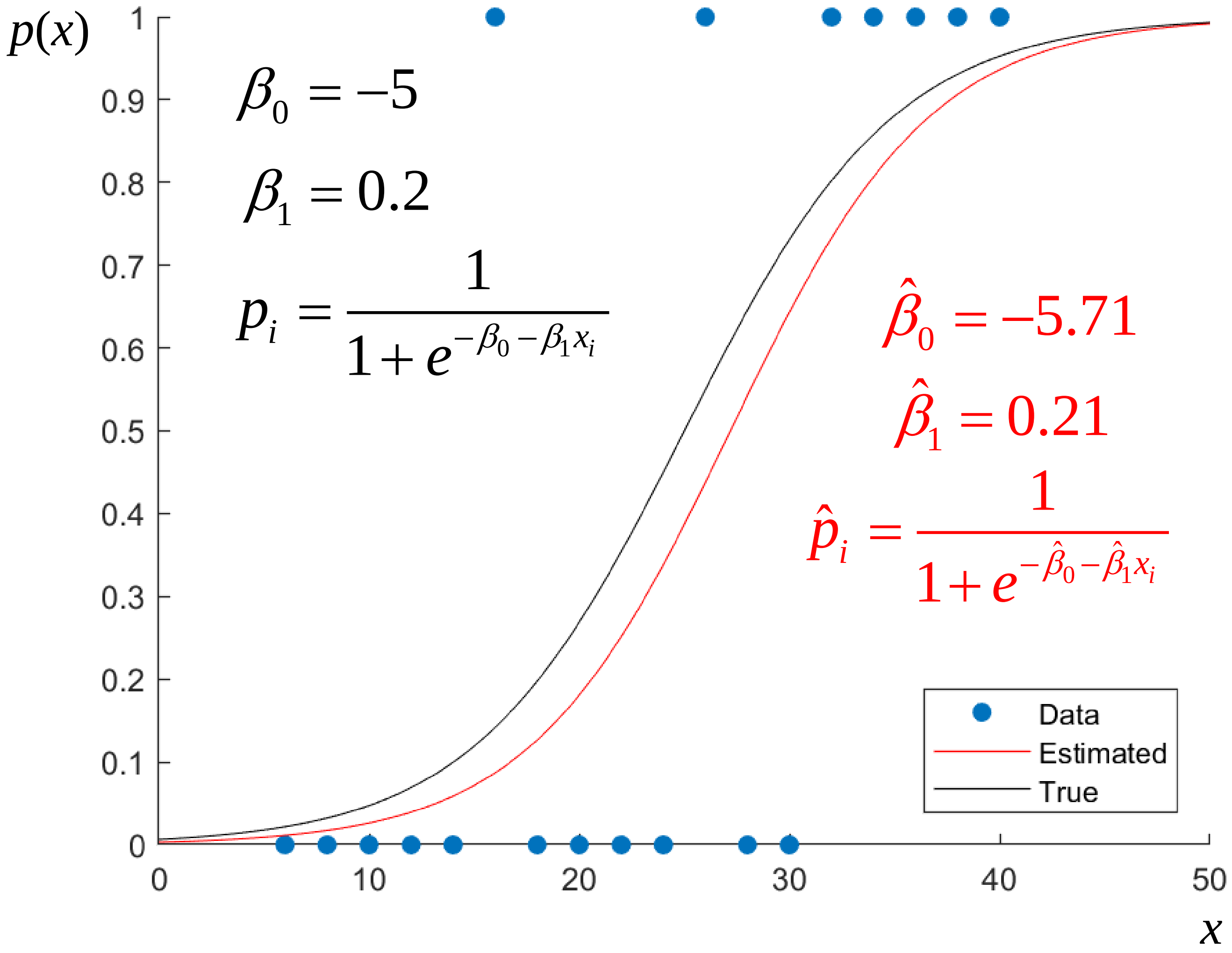
```
% estimated coefficient values,
proportions, and odds
[beta0hat,beta1hat]
LLhat=sum(y.*(beta0hat+beta1hat*x))...
-sum(log(1+exp(beta0hat+beta1hat*x)));
LLtru=sum(y.*(beta0tru+beta1tru*x))...
-sum(log(1+exp(beta0tru+beta1tru*x)));

% estimated proportions for succecss
Lhat=beta0hat+beta1hat*x;
phat=1./(1+exp(-Lhat));
[p,phat]
% odds
ohat=phat./(1-phat);
% proportions, and odds
[phat,ohat]
% OR
OR=exp(beta1hat*(24-22))
```

```
% likelihood surface
[X,Y]=meshgrid(beta1,beta0);
figure;
surf(X,Y,LL,'FaceAlpha',.8)
shading interp
hold on
scatter3(beta1hat,beta0hat,LLhat,25,r','filled')
scatter3(beta1tru,beta0tru,LLtru,25,'k','filled')
xlim([beta1min beta1max]),ylim([beta0min beta0max])
xlabel('beta0'), ylabel('beta1'), zlabel('Log Likelihood')
xhat=(0:.1:50)';
fxhat=1./(1+exp(-beta0hat-beta1hat.*xhat));
fxtru=1./(1+exp(-beta0tru-beta1tru.*xhat));
figure;
scatter(x,y,'filled')
hold on
plot(xhat,fxhat,'r')
plot(xhat,fxtru,'k')
legend('Data','Estimated','True','Location','SouthEast')
[beta0tru,beta0hat,beta1tru,beta1hat]
```

Logistic Regression Example





Logistic Regression Example

Once we have $\hat{\beta}_0$ and $\hat{\beta}_1$, insert them back into

$$\hat{p}_i = \frac{1}{1+e^{-\hat{\beta}_0 - \hat{\beta}_1 x_i}}$$ for estimated probabilities

and also for $\hat{o}_i = \dfrac{\hat{p}_i}{1-\hat{p}_i}$ odds

and for $\hat{OR} = e^{\hat{\beta}_1 \Delta}$, $\Delta = x_b - x_a$ odds ratio.

OR for a difference in x variable

$$\hat{\beta}_0 = -5.71$$
$$\hat{\beta}_1 = 0.21$$

$$\hat{OR} = e^{(0.21)(2)} = 1.5220$$

Hours (x)	A (y)	$\hat{p}$	$\hat{o}$
6	0	0.0115	0.0117
8	0	0.0175	0.0178
10	0	0.0263	0.0271
12	0	0.0395	0.0412
14	0	0.059	0.0627
16	1	0.0871	0.0954
18	0	0.1268	0.1451
20	0	0.1809	0.2209
22	0	0.2516	0.3362
24	0	0.3385	0.5117
26	1	0.4378	0.7788
28	0	0.5424	1.1853
30	0	0.6434	1.8040
32	1	0.7330	2.7456
34	1	0.8069	4.1787
36	1	0.8641	6.3598
38	1	0.9064	9.6794
40	1	0.9364	14.7317

Logistic Regression Discussion

The logistic regression model can be extended to more x's

$$p(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+...+\beta_q x_q)}} \quad , \quad -\infty < x_j < \infty, \; 0 \le p(x) \le 1$$

with

$$LL = \sum_{i=1}^{n} y_i(\beta_0 + \beta_1 x_{i1} + ... + \beta_q x_{iq}) - \sum_{i=1}^{n} \ln(1+e^{\beta_0+\beta_1 x_{i1}+...+\beta_q x_{iq}})$$

called Multiple Logistic Regression and to multiple outcomes,

Multinomial Logistic Regression AKA to ML people as Softmax.

Discussion

Questions?

Homework 8

1. Assume β_0=-10.0, β_1=0.3. Use $x=(10:5:60)'$.
   Generate a simulated data set.
   Estimate β_0, β_1, and p_i's by grid search and gradient descent.
   On one graph, plot graph (x_i,y_i) data, (x_i,p_i) and $(x_i,\hat{p}_i)$.
   Make a surface plot and show your estimated and true values.

Homework 8

2. Repeat 1. a large number of times.
   Make histograms of estimated $\hat{\beta}_0$ and $\hat{\beta}_1$.
   Calculate the sample means $\bar{\hat{\beta}}_0$ and $\bar{\hat{\beta}}_1$ along with standard deviations $s^2_{\hat{\beta}_0}$ and $s^2_{\hat{\beta}_1}$.

   On one graph plot points (x_i,y_i) data, $(x_i,p(\hat{\beta}_0 , \hat{\beta}_1))$ curve and some form of confidence interval that depends on $\hat{\beta}_0$, $s_{\hat{\beta}_0}$, $\hat{\beta}_1$, $s_{\hat{\beta}_1}$.

Homework 8

3. Obtain your own 0/1 data where p depends on x, $p(x)$.
   Fit a (multiple) logistic regression model to your data.
   Estimate β_0, β_1, and p_i's by grid search and gradient descent.
   On one graph, plot graph (x_i,y_i) data, (x_i,p_i) and $(x_i,\hat{p}_i)$.
   Make a surface plot and show your estimated values.

   Thoughts?