Data-Driven Decision-Making
In Enterprise Applications

Linear & Logistic Regression

Rainer Schlosser

Hasso Plattner Institute (EPIC)

May 2, 2019

Outline

- Questions regarding last Lecture?
- Problem Classifications & Solvers
- Today: Basic Regression Models
- Linear & Logistic Regression
- Homework

Problem Classifications & Solvers

Linear Continuous:	basically all solvers	
Linear Integer:	Cplex, Gurobi (+),	Minos (–)
Nonlinear Continuous:	Minos (+),	Cplex, Gurobi (–)
Nonlinear Integer:	Bonmin, Baron (+)	most solvers (–)

- Use linearizations and/or continuous relaxations to avoid Nonlinear Integer problems
- Use: `option solver './cplex';` or `option solver './minos';`

Linear Regression

Example: High Jump

- High Jump Results
- How they can be explained?   What are the key factors?
- Data: Results and features of participants (observations)
- What is a suitable regression model?
- How does it work?   What is the idea?
- How can we derive forecasts?
- How good are our forecasts?   Is there a measure?

High Jump Data

ID	Name	Result	Size	Gender	Party
1	Keven	160	176	1	0
2	Martin	155	178	1	0
3	Christian	140	172	1	1
4	Matthias	150	175	1	0
5	Ralf	130	160	1	0
6	Stefan	165	190	1	1
7	Markus	165	185	1	0
8	Cindy	130	168	0	0
9	Julia	130	163	0	1
10	Anna	145	170	0	0
11	Viktoria	155	171	0	0
12	Marilena	125	167	0	0

Notations

- Number of observations N in the example?
- Which quantity do we want to explain? (dependent variable y)
- Which quantities may be factors? (explanatory variables x)
- What might be missing variables?
- Mean of the dependent variable? $\overline{y} = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i$
- Variance of the dependent variable? $VAR = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \overline{y})^2$
- Plausibility checks: Expectations? Hypotheses?
- How do we quantify the impact/dependencies?

Least Squares Regression

- Idea: Use explanatory variables x to explain dependent variable y.
- Approach: Try to reconstruct y by linear parts of x

$$y_i \approx \beta_1 \cdot \underbrace{x_i^{(1)}}_{:=1} + \beta_2 \cdot x_i^{(2)} + \beta_3 \cdot x_i^{(3)} + \ldots \quad \text{with given data } \vec{x}_i, y_i, \; i = 1,..,N$$

β_k - coefficients have to be chosen such that the fit is “good”.

- What is a “good” fit? We need a measure.
- Answer: Minimize, e.g., the sum of squared deviations, i.e.,

$$\min_{\beta_1,\beta_2,\beta_3,\beta_4 \in \mathbb{R}} \sum_{i=1}^{N} \left(\beta_1 + \beta_2 \cdot x_i^{(2)} + \beta_3 \cdot x_i^{(3)} + \beta_4 \cdot x_i^{(4)} - y_i \right)^2$$

Solution & Forecasts

- We obtain optimal coefficients $\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$ (via a quadratic solver)

- What can we do with the coefficients $\vec{\beta}^* = (-102,\ 1.43,\ 3.05,\ -5.43)$?

- (1) We can quantify the impact of factors $x^{(2)}, x^{(3)}, x^{(4)}$ on y!

- (2) We can compute smart forecasts!

- Example: We have a new participant (179 tall, male, party: yes)

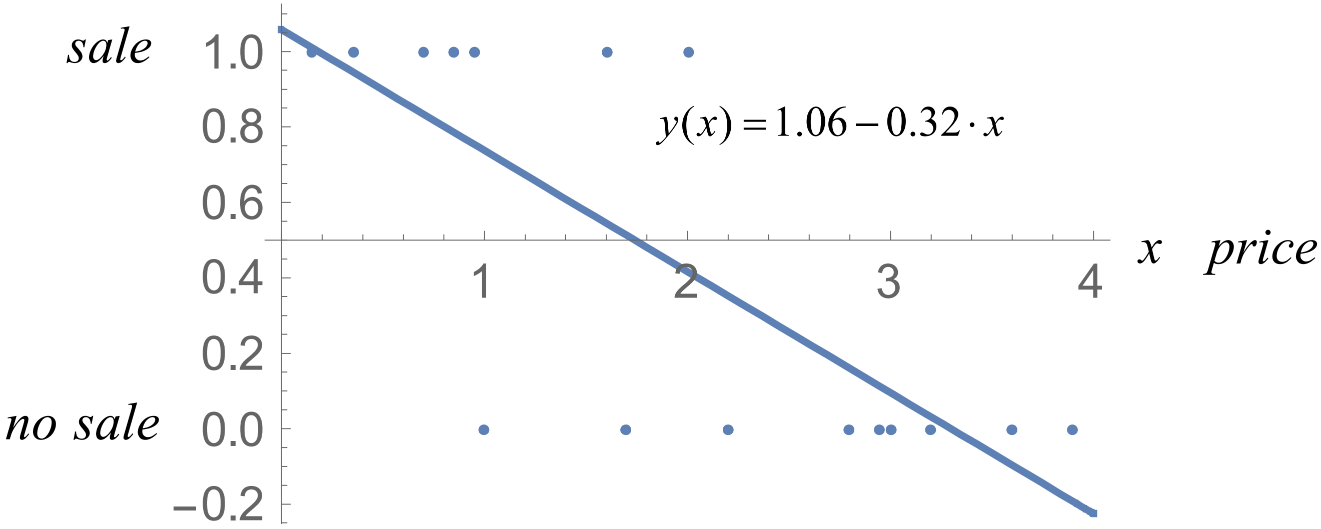
- Forecast: Estimated/expected result = $\beta_1^* + 179 \cdot \beta_2^* + 0 \cdot \beta_3^* + 1 \cdot \beta_4^* = 151.74$

How reliable is our Model?

- We can use various combinations of explanatory variables.
- We will always obtain a result and some optimal β^* coefficients!
- How to measure the quality of a model? There is a measure: R^2.
- Idea: How much of the variance in y can be explained by the model.
- Model fit: $\hat{y}_i = \beta_1^* + \beta_2^* \cdot x_i^{(2)} + \beta_3^* \cdot x_i^{(3)} + \ldots \approx y_i$
- New variance: $VAR_{new} = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \leq VAR = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \overline{y})^2$
- Goodness of fit: $R^2 = 1 - \frac{VAR_{new}}{VAR} \in [0,1]$ (large is good)

Logistic Regression

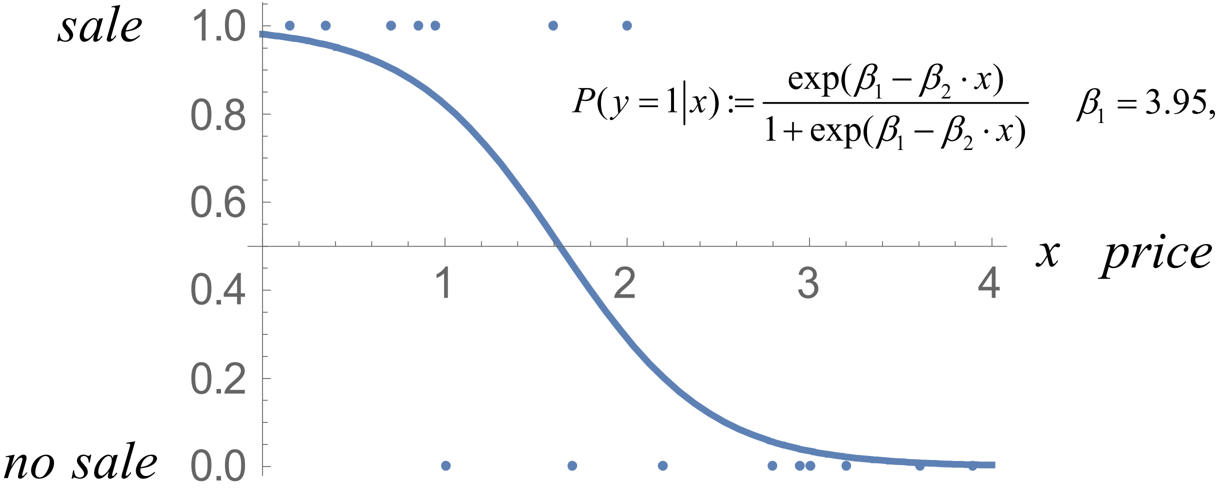
Estimation of Probabilities



Can the relation/prediction $y(x) = 1.06 - 0.32 \cdot x$ be used as sales probability?

Second Approach: Logistic Regression

- Binary 0/1 y observations, explanatory variable x, and **probabilities** $P(x)$



$$P(y=1|x) := \frac{\exp(\beta_1 - \beta_2 \cdot x)}{1+\exp(\beta_1 - \beta_2 \cdot x)} \qquad \beta_1 = 3.95,\ \ \beta_1 = 2.42$$

- What is the idea behind logistic regression?

Approach: Maximum Likelihood Estimation

- Idea: (1) Choose a model + (2) Find the best calibration
- Toy Example: Coin Toss
- Data: 010111010100010001010010001100000
- Model: Bernoulli Experiment with success probability *p*
- Calibration: Which model, i.e., which *p* explains our data best?

Our Model: Bernoulli Distribution

- Random variable Y sale occurred (1 yes, 0 no)

- Success probability $P(Y=1)=p$ and $P(Y=0)=1-p$

- Bernoulli distribution $P(Y=k)=p^k\cdot(1-p)^{1-k}, \quad k=0,1$

- (Binomial distribution) $P(Y=k)=\binom{n}{k}\cdot p^k\cdot(1-p)^{n-k},$

  for multiple sales $k=0,....,n$ (cf. n=1)

Likelihood Function

- Bernoulli distribution $P(Y=k)=p^k\cdot(1-p)^{1-k},\quad k=0,1$
- Consider observed data $\vec{y}=(y_1,\ldots,y_N),\quad y_i\in\{0,1\},\quad i=1,\ldots,N$
- Probability for one obs. $P(Y_i=y_i)=p^{y_i}\cdot(1-p)^{1-y_i},\quad y_i\in\{0,1\}$
- *Joint probability* $$P(Y_1=y_1,\ldots,Y_N=y_N)=\prod_{i=1}^{N}P(Y_i=y_i)$$

  (Likelihood Function) $$=\prod_{i=1}^{N}p^{y_i}\cdot(1-p)^{1-y_i}$$
- Now, maximize the joint probability over the success probability p!

Maximize the Likelihood Function

$$\max P(Y_1 = y_1, \ldots, Y_N = y_N)$$ i.i.d. (independent, identically distributed)

$$= \max_{p} \prod_{i=1}^{N} P(Y_i = y_i)$$

$$= \max_{p \in [0,1]} \prod_{i=1}^{N} p^{y_i} \cdot (1-p)^{1-y_i}$$

Actually, we wanted to find the best *p*.

$$\arg\max_{p \in [0,1]} \prod_{i=1}^{N} p^{y_i} \cdot (1-p)^{1-y_i}$$

We are interested in First Order Conditions. Hence, we do not like products!

Monotone Increasing Transformations

$$\underset{p\in[0,1]}{\arg\max}\left\{\prod_{i=1}^{N} p^{y_i}\cdot(1-p)^{1-y_i}\right\}$$

$$=\underset{p\in[0,1]}{\arg\max}\left\{5\cdot\left(\prod_{i=1}^{N} p^{y_i}\cdot(1-p)^{1-y_i}\right)+17\right\}$$? (linear)

$$=\underset{p\in[0,1]}{\arg\max}\left\{\left(\prod_{i=1}^{N} p^{y_i}\cdot(1-p)^{1-y_i}\right)^2\right\}$$?? (convex)

$$=\underset{p\in[0,1]}{\arg\max}\left\{\ln\left(\prod_{i=1}^{N} p^{y_i}\cdot(1-p)^{1-y_i}\right)\right\}$$??? (concave)

Log-Likelihood Function

$$\arg\max_{p} P(Y_1 = y_1, \ldots, Y_N = y_N)$$

$$= \arg\max_{p \in [0,1]} \left\{ \ln\left(\prod_{i=1}^{N} p^{y_i} \cdot (1-p)^{1-y_i} \right) \right\}$$

$$= \arg\max_{p \in [0,1]} \left\{ \sum_{i=1}^{N} \ln\left(p^{y_i} \cdot (1-p)^{1-y_i} \right) \right\}$$

$$= \arg\max_{p \in [0,1]} \left\{ \sum_{i=1}^{N} \left(\ln\left(p^{y_i} \right) + \ln\left((1-p)^{1-y_i} \right) \right) \right\}$$

$$= \arg\max_{p \in [0,1]} \left\{ \sum_{i=1}^{N} \left(y_i \cdot \ln(p) + (1-y_i) \cdot \ln(1-p) \right) \right\}$$

Optimization

- FOC: $$\frac{\partial}{\partial p} P(Y_1 = y_1, \dots, Y_N = y_N) \overset{!}{=} 0$$

$$\sum_{i=1}^{N} \left(y_i \cdot \ln(p)' + (1 - y_i) \cdot \ln(1-p)' \right) \overset{!}{=} 0$$

$$\Leftrightarrow \quad \sum_{i=1}^{N} \left(\frac{y_i}{p} + \frac{1 - y_i}{1-p} \right) \overset{!}{=} 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N} \left(\underbrace{(1-p) \cdot y_i + p \cdot (1 - y_i)}_{= y_i + (1 - 2 y_i) \cdot p} \right) \overset{!}{=} 0$$

- Solve for *p*.
- 1 Variable, 1 Equation (Unique solution p^*)
- **Result**: Our data fits to the model $P(Y=1) = p^*$ and $P(Y=0) = 1 - p^*$.

Generalization & Pricing Use Case

Use Case: Demand Estimation on Amazon

- Regular price adjustments (e.g., time intervals of ca. 2 hours)

- Observation of market conditions (at the time of price adjustments)

  e.g., Competitors' prices, quality, rating, shipping time, etc.

- Sales observations: Points in time (within certain intervals)

- Rare events, i.e., 0 or 1 sales between price adjustments (2 hours)

A Seller’s Data Set

period	**sale**	**price**	rank	competitor’s prices for product *i* (ISBN)				
t	$y_t^{(i)}$	$a_t^{(i)}$	$r_t^{(i)}$	$p_{t,1}^{(i)}$	$p_{t,2}^{(i)}$	$p_{t,3}^{(i)}$	$p_{t,4}^{(i)}$	… $p_{t,K}^{(i)}$
1	**0**	**19**	3	13	17	20	25	
2	**0**	**15**	2	13	17	20	25	
3	**1**	**10**	1	13	15	20	/	
4	**0**	**10**	1	13	15	20	22	
5	**1**	**12**	2	11	15	20	24	
6	**0**	**15**	3	11	14	20	24	
…								

Estimation of Sales Probabilities

- Goal: Quantify sales probabilities as function of our offer price
- Idea: Sales probabilities should depend on market conditions
- Approach: Maximum Likelihood

(1) Choose family of models: Logistic function

(2) Define explanatory variables (based on our data)

(3) Calibrate model: Find model coefficients

(4) Result: Quantify sales probabilities for any market situation!

Explanatory Variables

- Data: Market situation in *t*: $\vec{s} = (t, p_1, \ldots, p_K, q_1, \ldots, p_K, r_1, \ldots, r_K, f_1, \ldots, f_K, \ldots)$

- Define explanatory variables (What could affect decisions?):

$x_1(a, \vec{s}) := 1$	(Intercept)
$x_2(a, \vec{s}) := \mathit{price\ rank}$	(Rank of offer price within competitors' prices)
$x_3(a, \vec{s}) := a - \min\limits_{k=1,\ldots,K} p_k$	(Price difference to best competitor)
$x_4(a, \vec{s}) := \mathit{quality\ rank}$	(Rank of our product condition)
$x_5(a, \vec{s}) := \#\mathit{commercials}$	(Number of competitors with feedback >10000)
$x_6(a, \vec{s}) := \mathit{combinations}$	(Number of comp. with better price + better quality)
$x_7(a, \vec{s}) := 1_{\{a \cdot 100 \mod 10 = 9\}}$	(Psychological Prices)

. . .

One Family of Models: Logistic Function

- $P\big(Y=1 \mid \vec{x}(a,\vec{s})\big) := e^{\vec{x}'\vec{\beta}} / (1+e^{\vec{x}'\vec{\beta}})$

$$= \frac{\exp\big(\beta_1 \cdot x_1(a,\vec{s}) + \beta_2 \cdot x_2(a,\vec{s}) + \ldots\big)}{1+\exp\big(\beta_1 \cdot x_1(a,\vec{s}) + \beta_2 \cdot x_2(a,\vec{s}) + \ldots\big)} \in (0,1)$$

- There are other families, but this is a good family

- Maximum Likelihood Estimation:

  Find best $\vec{\beta}$ coefficients for our data $y_t, \vec{x}(a_t, \vec{s}_t)$, $t = 1, \ldots, N$

Maximize the Log-Likelihood Function

- Recall:

$$\arg\max_{p} P(Y_1 = y_1, \ldots, Y_N = y_N) = \arg\max_{p \in [0,1]} \left\{ \sum_{i=1}^{N} \left(y_i \cdot \ln(p) + (1 - y_i) \cdot \ln(1 - p) \right) \right\}$$

- Now, we have the conditional probabilities:

$$\arg\max_{\vec{\beta}} P\left(Y_1 = y_1 \mid a_1, \vec{s}_1, \ \ldots \ , Y_N = y_N \mid a_N, \vec{s}_N\right)$$

$$= \arg\max_{\beta_m \in \mathbb{R}, m=1,\ldots,M} \left\{ \sum_{i=1}^{N} \left(y_i \cdot \ln\left(\frac{e^{\vec{x}(a_i, \vec{s}_i)'\vec{\beta}}}{1 + e^{\vec{x}(a_i, \vec{s}_i)'\vec{\beta}}} \right) + (1 - y_i) \cdot \ln\left(1 - \frac{e^{\vec{x}(a_i, \vec{s}_i)'\vec{\beta}}}{1 + e^{\vec{x}(a_i, \vec{s}_i)'\vec{\beta}}} \right) \right) \right\}$$

Optimization

- FOC: $\frac{\partial}{\partial \vec{\beta}} P(Y_1 = y_1 \mid a_1, \vec{s}_1, \;\ldots\;, Y_N = y_N \mid a_N, \vec{s}_N) \overset{!}{=} 0$

$$\sum_{i=1}^{N} \left(y_i \cdot \frac{\partial}{\partial \beta_m} \ln\left(\frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}} \right) + (1-y_i) \cdot \frac{\partial}{\partial \beta_m} \ln\left(1 - \frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}} \right) \right) \overset{!}{=} 0, \quad m = 1, \ldots, M$$

$$\Leftrightarrow \sum_{i=1}^{N} \left(\left(y_i - \frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}} \right) \cdot x_i^{(m)} \right) \overset{!}{=} 0, \quad \text{for all } m = 1, \ldots, M$$

- Solve **the nonlinear system** for $\vec{\beta} = (\beta_1, \ldots, \beta_M)$
- M Variables, M Equations (Unique solution $\vec{\beta}^* = (\beta_M^*, \ldots, \beta_M^*)$)
- Result: Our data fits to the model $P\left(Y = 1 \mid \vec{x}(a,\vec{s})\right) := e^{\vec{x}(a,\vec{s})'\vec{\beta}^*} / (1 + e^{\vec{x}(a,\vec{s})'\vec{\beta}^*})$

Check Proof

$$\sum_{i=1}^{N}\left(y_i\cdot\left(\frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}\right)^{-1}\frac{\partial}{\partial\beta_m}\left(1+e^{-\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}\right)^{-1}+(1-y_i)\cdot\frac{\partial}{\partial\beta_m}\ln\left(1-\frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}\right)\right)\overset{!}{=}0$$

$$\sum_{i=1}^{N}\left(y_i\cdot\left(\frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}\right)^{-1}(-1)\left(1+e^{-\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}\right)^{-2}\cdot e^{-\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}(-x_i^{(k)})-(1-y_i)\cdot\left(\frac{1}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}\right)^{-1}\left(1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}\right)^{-2}\cdot e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}(x_i^{(k)})\right)\overset{!}{=}0$$

$$\sum_{i=1}^{N}\left(y_i\cdot\left(1+e^{-\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}\right)^{1}(-1)\left(1+e^{-\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}\right)^{-2}\cdot e^{-\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}(-x_i^{(k)})-(1-y_i)\cdot\left(1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}\right)^{-1}\cdot e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}(x_i^{(k)})\right)\overset{!}{=}0$$

$$\sum_{i=1}^{N}\frac{1}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}\cdot\left(y_i\cdot(x_i^{(k)})-(1-y_i)\cdot(x_i^{(k)})\right)\frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}\overset{!}{=}0$$

$$\sum_{i=1}^{N}\left(y_i\cdot(x_i^{(k)})-(x_i^{(k)})\frac{e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}{1+e^{\vec{x}(a_i,\vec{s}_i)'\vec{\beta}}}\right)\overset{!}{=}0$$

Application of the Model Obtained

- Observe current market situation for a product: $\vec{s}$
- For *any* admissible offer prices a we can evaluate $\vec{x}(a,\vec{s})$ and obtain

$$P\left(Y=1\,|\,\vec{x}(a,\vec{s})\right):=\frac{e^{\vec{x}(a,\vec{s})'\vec{\beta}^*}}{1+e^{\vec{x}(a,\vec{s})'\vec{\beta}^*}}$$

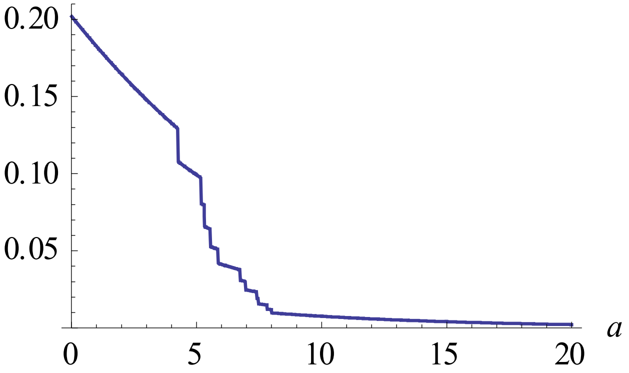
- Now, we can optimize *expected profits* (for one time interval):

$$\max_{a\geq 0}\left\{(a-c)\cdot\frac{e^{\vec{x}(a,\vec{s})'\vec{\beta}^*}}{1+e^{\vec{x}(a,\vec{s})'\vec{\beta}^*}}\right\}$$

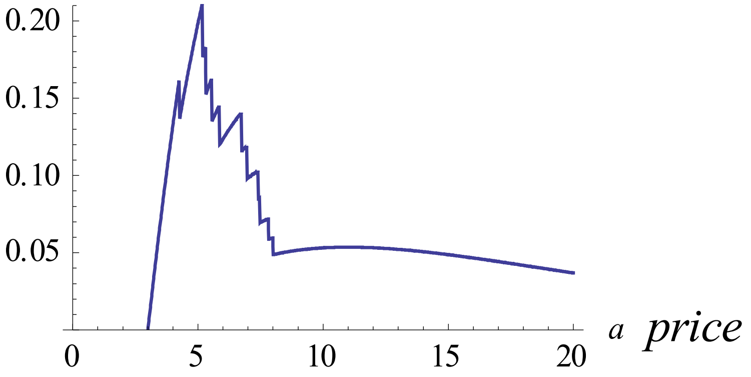
Prediction of Sales Probabilities

- Example: Competitor’s prices $\vec{p} = (4.26,\ 5.18,\ 5.31,\ 5.55,\ 5.86,\ \ldots)$

sales probability $P(a)$



expected profit $(a-3)\cdot P(a)$



Summary

(+) Logistic Regression is simple and robust

(+) Allows for many observations *N* and many features *M*

(+) Plausibility Checks & Closed Form Expressions

(+/–) Definition of Customized Explanatory Variables

(–) No dependencies between variables

(–) Limited to binary dependent variables

What is a good Model?

- Use "Goodness of fit" measures (for MLE models)

- *AIC* (low is good, trade-of between fit and number of variables *M*)

  $$AIC := -2 \cdot \sum_{i=1}^{N} \left(y_i \cdot \ln p_i + (1-y_i) \cdot \ln(1-p_i) \right) + 2 \cdot M$$

  Note, p_i depends on all features x_i and the optimal β^* coefficients.

- Normalized (McFadden Pseudo R^2): $R^2 := 1 - AIC / AIC_0$ (vs. Null-model)

- Be creative: Test different variables and find the smallest *AIC* value.

  Hint: Not quantity but quality counts!

Next Lecture (May 16)

Homework: Solve at least 2 out of 4 of the following problems

- The latest version of the HPI Master Project Assignment Problem
- A model to solve Sudoku examples
- Soccer line-up with several constraints
- A model to solve logistic regressions

Hand in (May 23): (i) written key formulas/model & (ii) executable (Ampl) files

Teams of at most two students are allowed. Questions per Mail.

For further details take a look at the course website.

Overview

2	April 25	Linear Programming I
3	April 29	Linear Programming II
4	May 2	Linear/Logistic Regression + Homework (3 weeks time)
5	**May 16**	**Exercise Implementations**
6	May 20	Dynamic Programming
7	May 23	Response Strategies / Game Theory
8	May 27	Project Assignments
9	June 3	Robust Optimization
10	June 13	Workshop / Group Meetings
11	June 20	Presentations (First Results)
12/13	June 24/27	Workshop / Group Meetings
14/15	July 1/4	Workshop / Group Meetings
16	July 11	Presentations (Final Results), Feedback, Documentation (Aug 31)