Machine Learning

Logistic Regression

Logistic regression

- Name is somewhat misleading. Really a technique for classification, not regression.
  - "Regression" comes from fact that we fit a linear model to the feature space.
- Involves a more probabilistic view of classification.

Different ways of expressing probability

- Consider a two-outcome probability space, where:
  - $p(O_1) = \mathrm{p}$
  - $p(O_2) = 1 - \mathrm{p} = \mathrm{q}$
- Can express probability of O_1 as:

	notation	range equivalents		
standard probability	p	0	0.5	1
odds	p / q	0	1	+ ∞
log odds (logit)	log(p / q)	- ∞	0	+ ∞

Log odds

- Numeric treatment of outcomes O_1 and O_2 is equivalent
  - If neither outcome is favored over the other, then log odds = 0.
  - If one outcome is favored with log odds = x, then other outcome is disfavored with log odds = $-x$.
- Especially useful in domains where relative probabilities can be miniscule
  - Example: multiple sequence alignment in computational biology

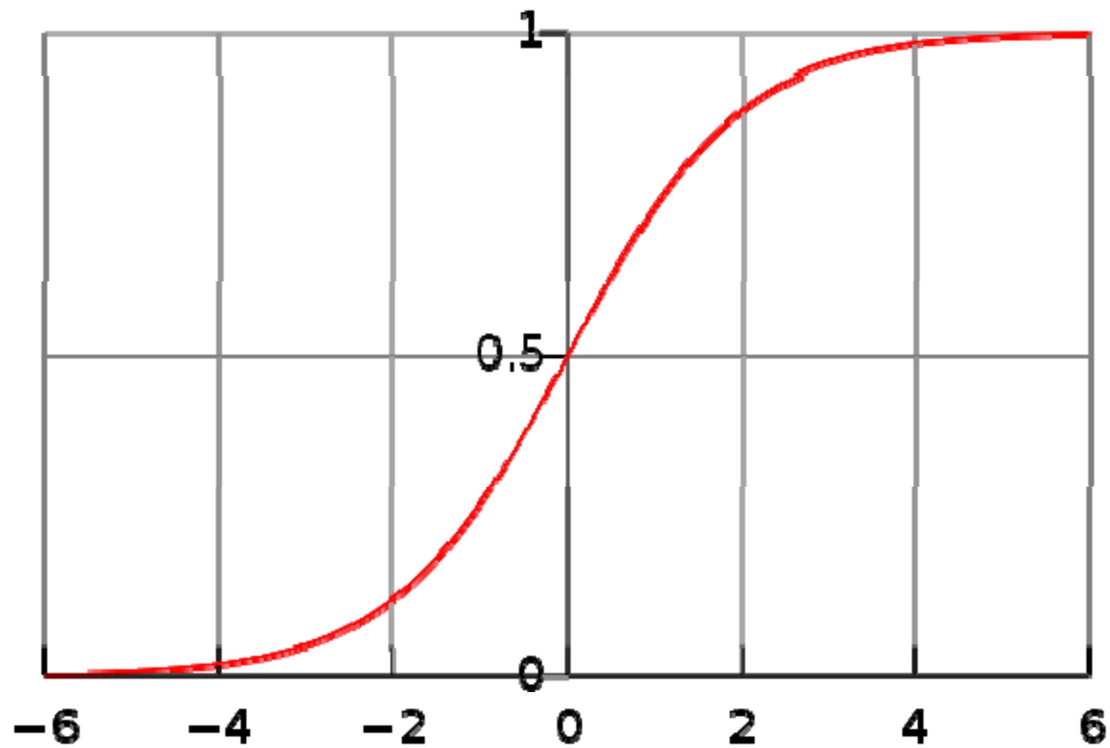
From probability to log odds (and back again)

$$z = \log\left(\frac{p}{1-p}\right)$$ logit function

$$\frac{p}{1-p} = e^z$$

$$p = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$ logistic function

Standard logistic function



Logistic regression

- Scenario:
  - A multidimensional feature space (features can be categorical or continuous).
  - Outcome is discrete, not continuous.
    - We'll focus on case of two classes.
  - It seems plausible that a linear decision boundary (hyperplane) will give good predictive accuracy.

Using a logistic regression model

- Model consists of a vector β in *d*-dimensional feature space

- For a point **x** in feature space, project it onto β to convert it into a real number *z* in the range - ∞ to + ∞

$$z = \alpha + \boldsymbol{\beta} \cdot \mathbf{x} = \alpha + \beta_1 x_1 + \cdots + \beta_d x_d$$

- Map *z* to the range 0 to 1 using the logistic function

$$p = 1/(1 + e^{-z})$$

- Overall, logistic regression maps a point **x** in *d*-dimensional feature space to a value in the range 0 to 1

Using a logistic regression model

- Can interpret prediction from a logistic regression model as:
  - A probability of class membership
  - A class assignment, by applying threshold to probability
    - threshold represents decision boundary in feature space

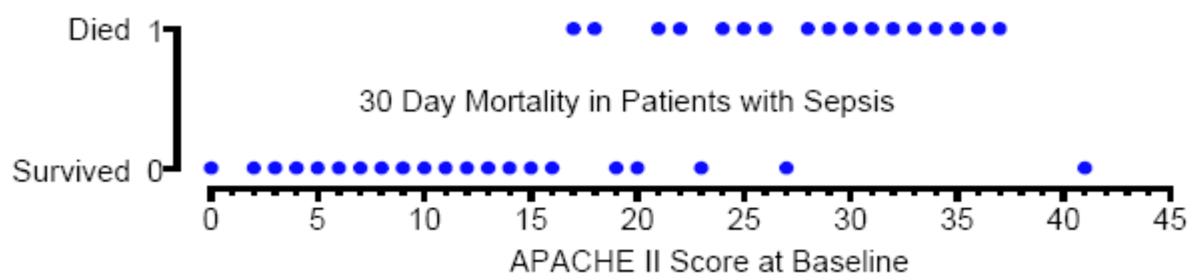
Training a logistic regression model

- Need to optimize β so the model gives the best possible reproduction of training set labels
  - Usually done by numerical approximation of maximum likelihood
  - On really large datasets, may use stochastic gradient descent

Logistic regression in one dimension

a) Example: APACHE II Score and Mortality in Sepsis

The following figure shows 30 day mortality in a sample of septic patients as a function of their baseline APACHE II Score. Patients are coded as 1 or 0 depending on whether they are dead or alive in 30 days, respectively.

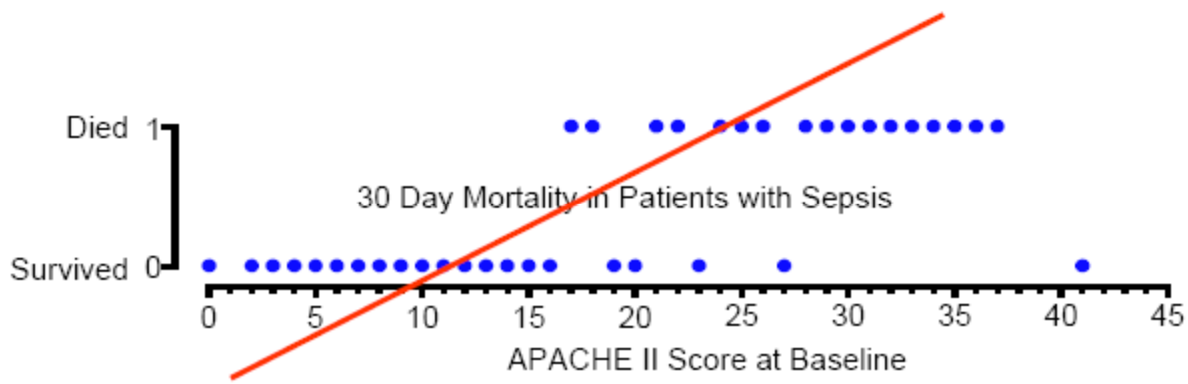


Logistic regression in one dimension

We wish to predict death from baseline APACHE II score in these patients.

Let $\pi(x)$ be the probability that a patient with score x will die.

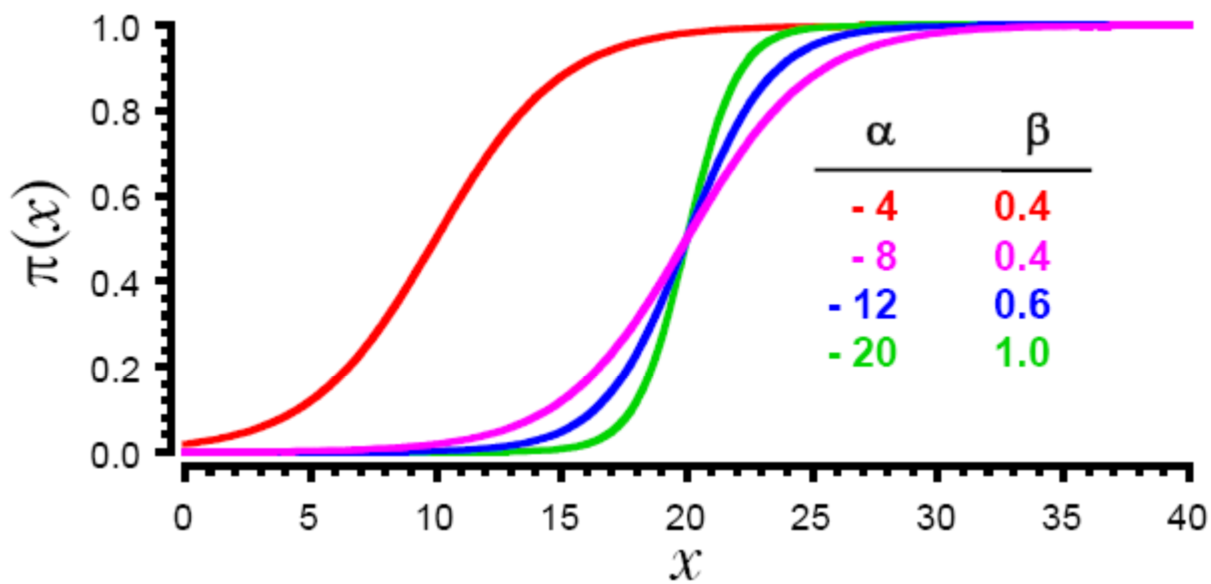
Note that linear regression would not work well here since it could produce probabilities less than zero or greater than one.



Logistic regression in one dimension

- Parameters control shape and location of sigmoid curve
  - α controls location of midpoint
  - β controls slope of rise

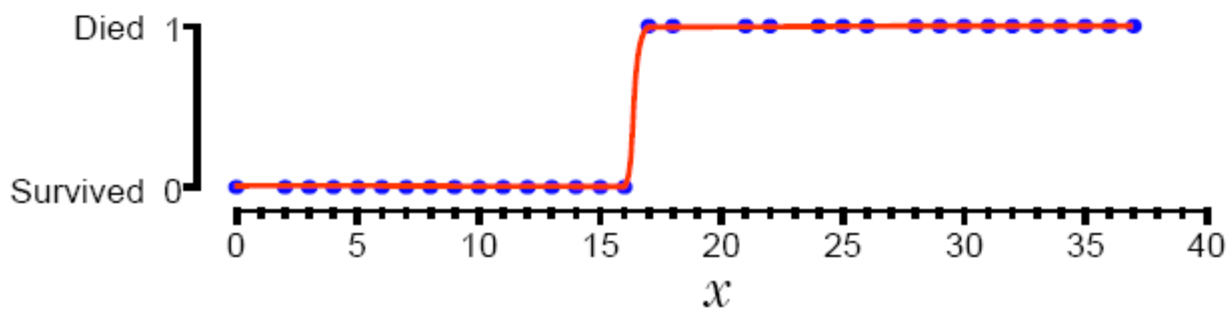
$$\pi(x) = \exp(\alpha + \beta x) / (1 + \exp(\alpha + \beta x))$$



When $x = -\alpha / \beta$, $\alpha + \beta x = 0$ and hence $\pi(x) = 1/(1+1) = 0.5$

Logistic regression in one dimension

Data that has a sharp survival cut off point between patients who live or die should have a large value of β.



Data with a lengthy transition from survival to death should have a low value of β.