Logistic Regression

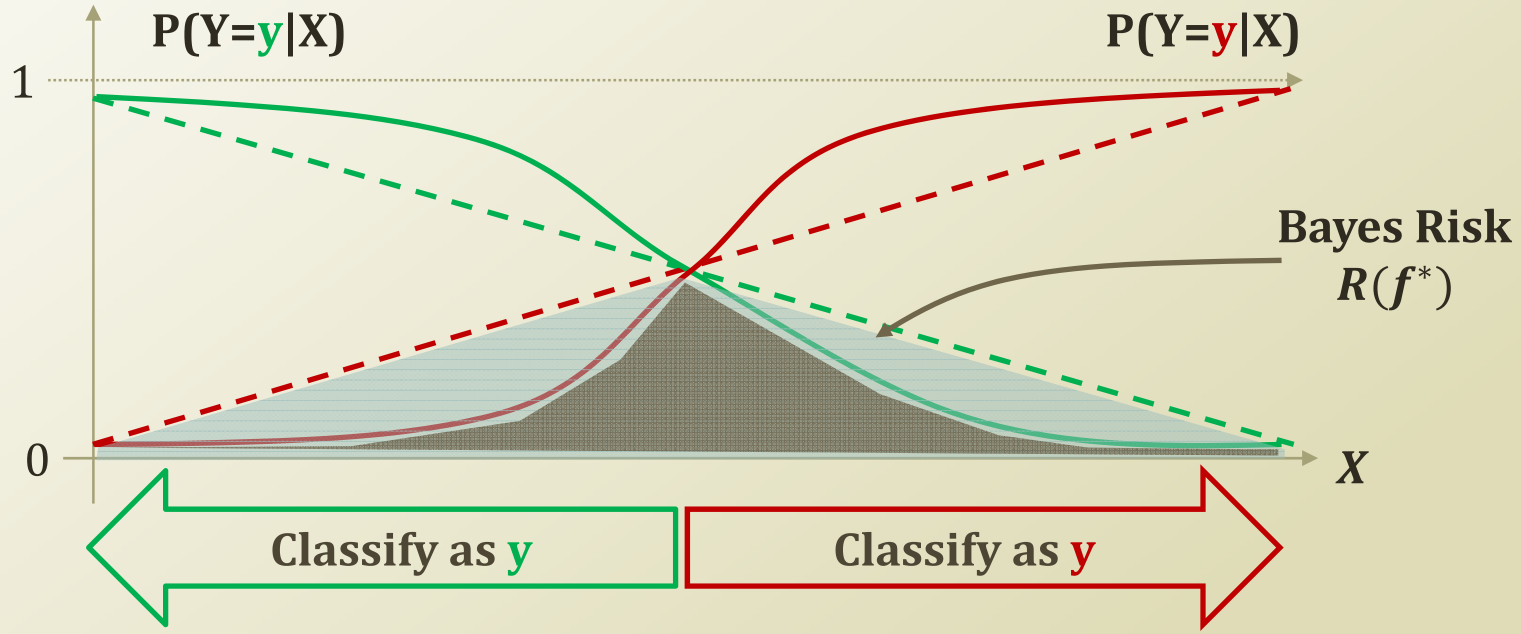
Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

- Learn the logistic regression classifier
  - Understand why the logistic regression is better suited than the linear regression for classification tasks
  - Understand the logistic function
  - Understand the logistic regression classifier
  - Understand the approximation approach for the open form solutions
- Learn the gradient descent algorithm
  - Know the tailor expansion
  - Understand the gradient descent/ascent algorithm
- Learn the different between the naïve Bayes and the logistic regression
  - Understand the similarity of the two classifiers
  - Understand the differences of the two classifiers
  - Understand the performance differences

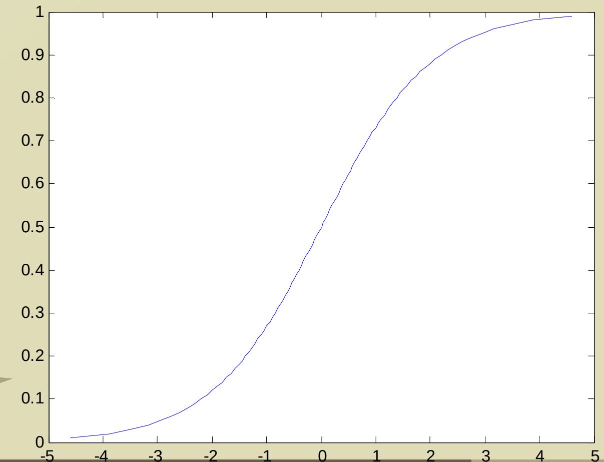
LOGISTIC REGRESSION

Optimal Classification and Bayes Risk



- Linear function vs. Non-linear function of P(Y|X)
  - Which is better?
- Problems of linear function
  - Range
  - Risk optimization
- Which function to use?
  - Need S-curve!

S-curve
a.k.a. Sigmoid
function



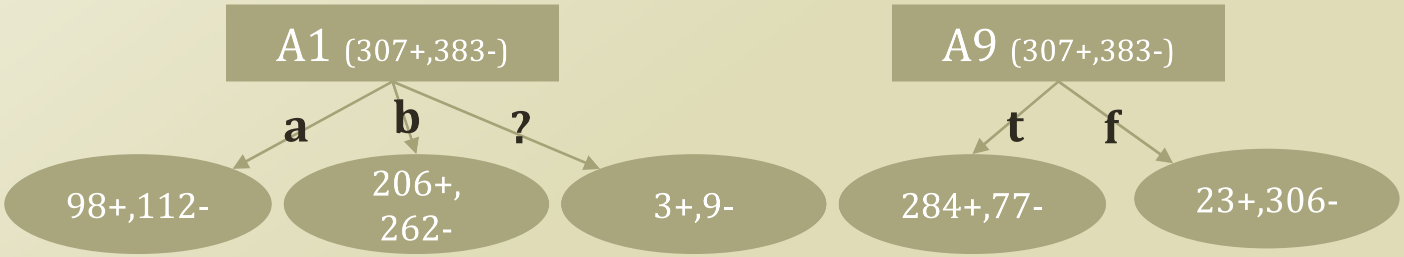
KAIST

Detour: Credit Approval Dataset

- http://archive.ics.uci.edu/ml/datasets/Credit+Approval
- To protect the confidential information, the dataset is anonymized
  - Feature names and values, as well
- A1: b, a.
  A2: continuous.
  A3: continuous.
  A4: u, y, l, t.
  A5: g, p, gg.
  A6: c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff.
  A7: v, h, bb, j, n, z, dd, ff, o.
  A8: continuous.
  A9: t, f.
  A10: t, f.
  A11: continuous.
  A12: t, f.
  A13: g, p, s.
  A14: continuous.
  A15: continuous.
  C: +,- (class attribute)

Some Counting Result

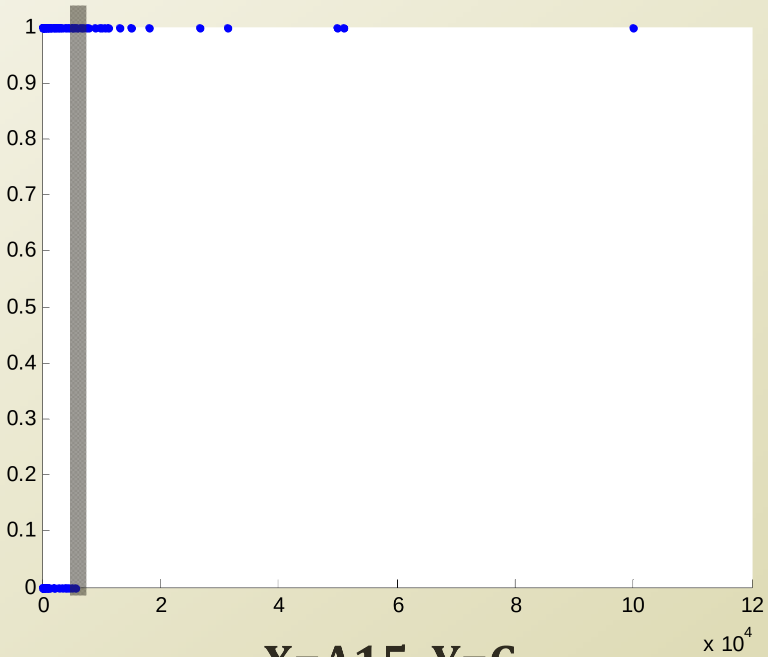
- 690 instances total
- 307 positive instances
- Considering A1
  - 98 positive when a
  - 112 negative when a
  - 206 positive when b
  - 262 negative when b
  - 3 positive when ?
  - 9 negative when ?
- Considering A9
  - 284 positive when t
  - 77 negative when t
  - 23 positive when f
  - 306 negative when f



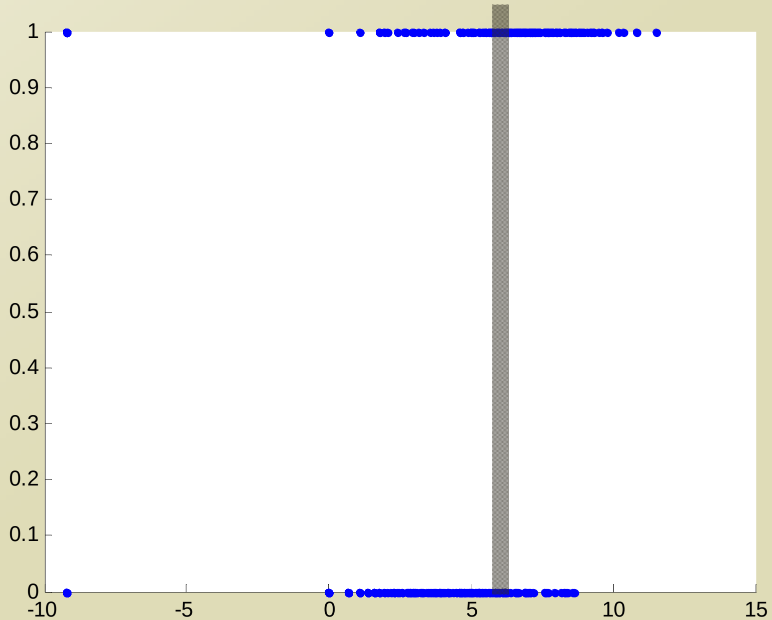
Which is a better attribute to include in the feature set of the hypothesis?

Classification with One Variable

- Let's predict the class, C, with an attribute, A15
  - Imagine that the Y axis shows P(Y|X)
  - There is a decision boundary
    - You can see it intuitively
- Then, How to find the boundary?



X=A15, Y=C

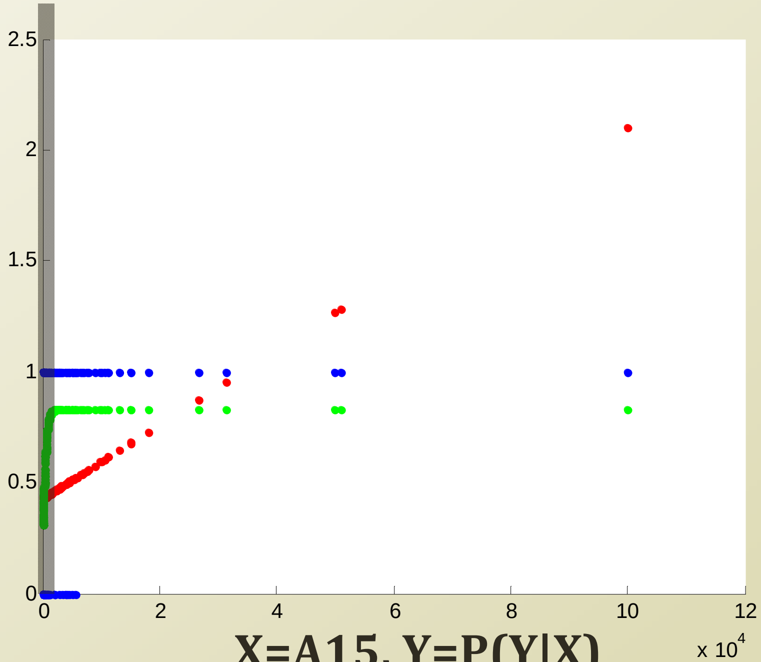


X=log(A15), Y=C

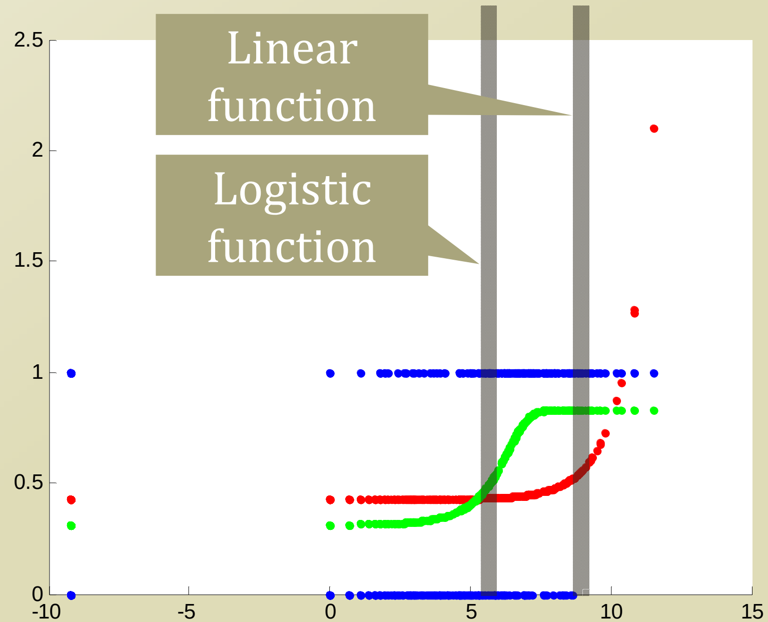
Linear Function vs. Non-Linear Function

- Problem of fitting to the linear function
  - Violate the probability axiom
  - Slow response to the examples
- Better to fit to the logistic function
  - Keep the probability axiom
  - Quick response around the decision boundary
- Which function to use?
  - Logistic function – a special case of sigmoid function

Blue = (X, Y_{true})
Red = $(X, P_{lin}(Y|X))$
Green = $(X, P_{log}(Y|X))$



X=A15, Y=P(Y|X)



X=log(A15), Y=P(Y|X)