

INTRO TO DATA SCIENCE

LECTURE 5: LOGISTIC REGRESSION

December 8, 2014

DAT11-SF

LAST TIME:

- LINEAR REGRESSION
- POLYNOMIAL REGRESSION
- REGULARIZATION

QUESTIONS?

I. LINEAR CLASSIFIERS (MODELS)**II. LOGISTIC REGRESSION****LAB:****III. LOGISTIC REGRESSION SCIKIT-LEARN**

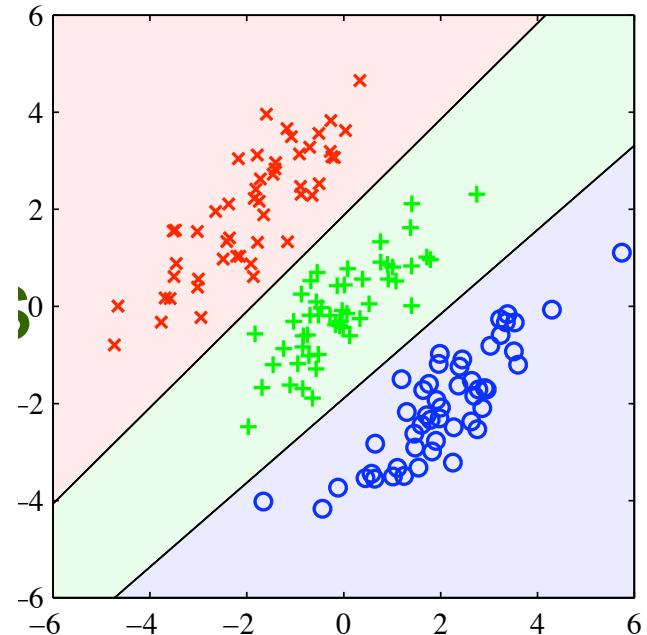
INTRO TO DATA SCIENCE

I. LINEAR MODELS

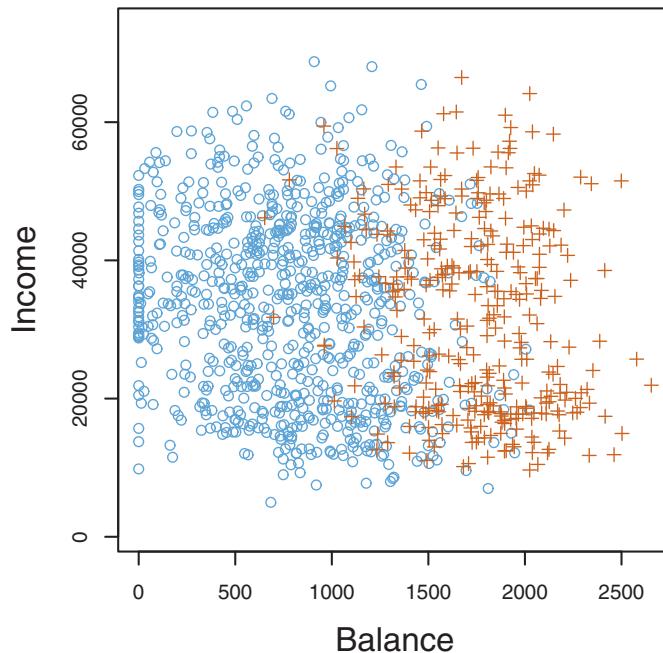
- Scoring function based on linear combination of features: $y = w_0 + \sum_{i=1}^d w_i x_i,$
- Decision boundaries also linear function of input:

- Input space divided into decision regions
- Decision surfaces (boundaries) are linear functions of input x
- D-1 dim. hyperplane within D dim. input space

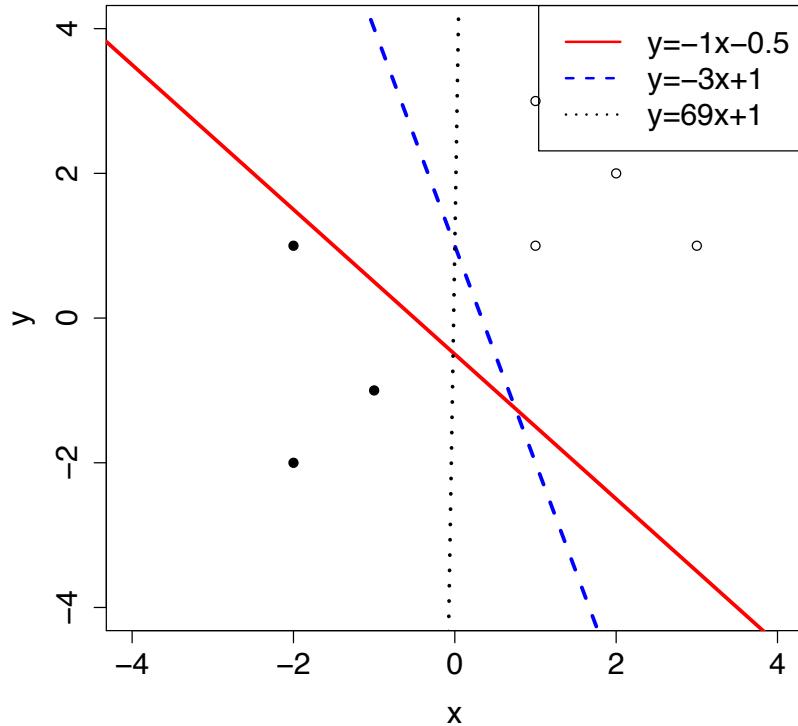
$$w_1x_1 + w_2x_2, \dots, +w_dx_d + w_0 = 0$$



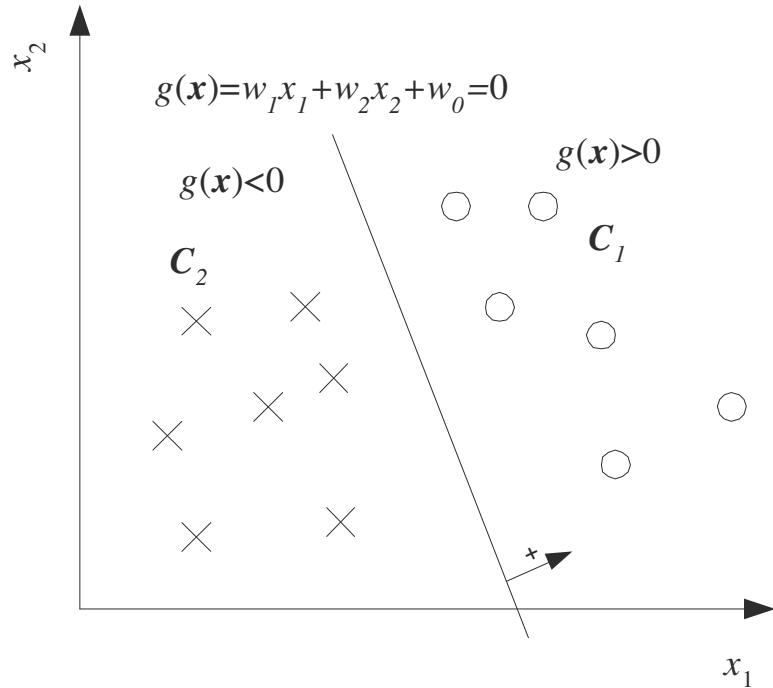
LINEAR CLASSIFICATION MODEL



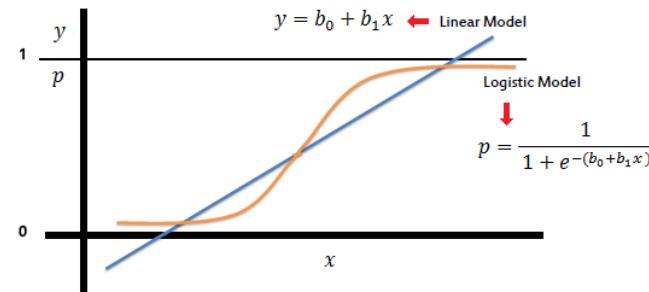
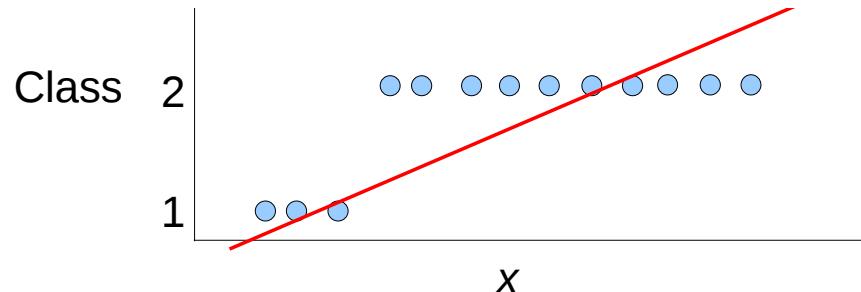
- Scoring function based on linear combination of features:
- Decision boundaries also linear function of input:



GEOMETRY OF 2D LINEAR MODEL



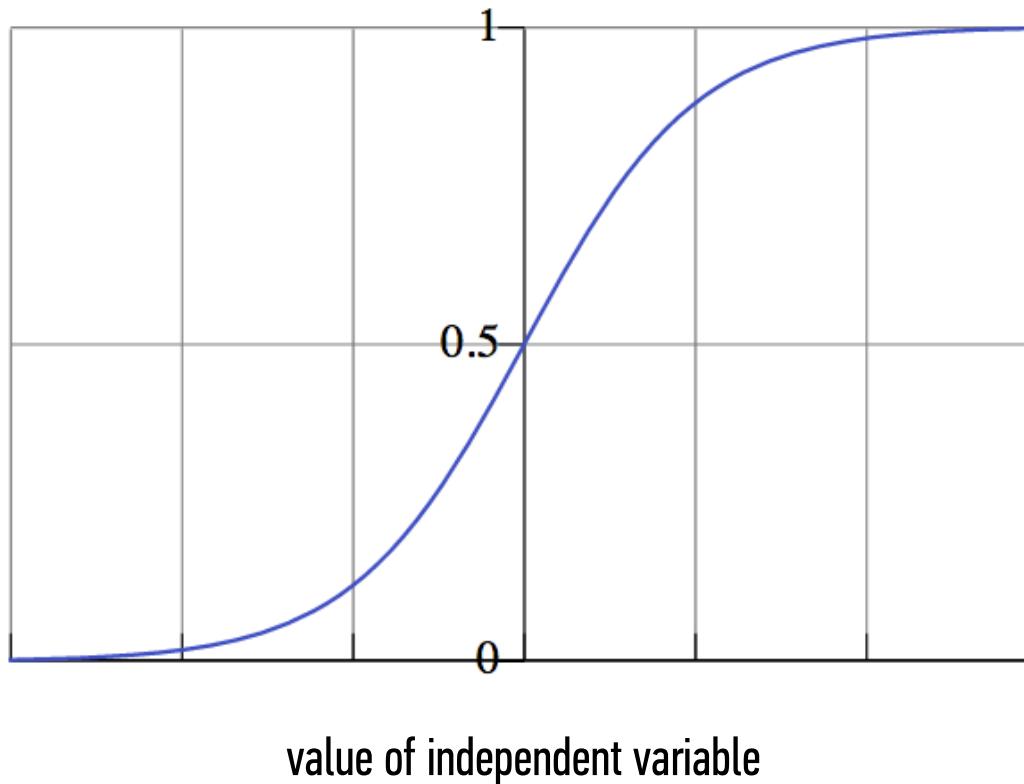
- Scoring function based on linear combination of features:
- Decision boundaries also linear function of input:



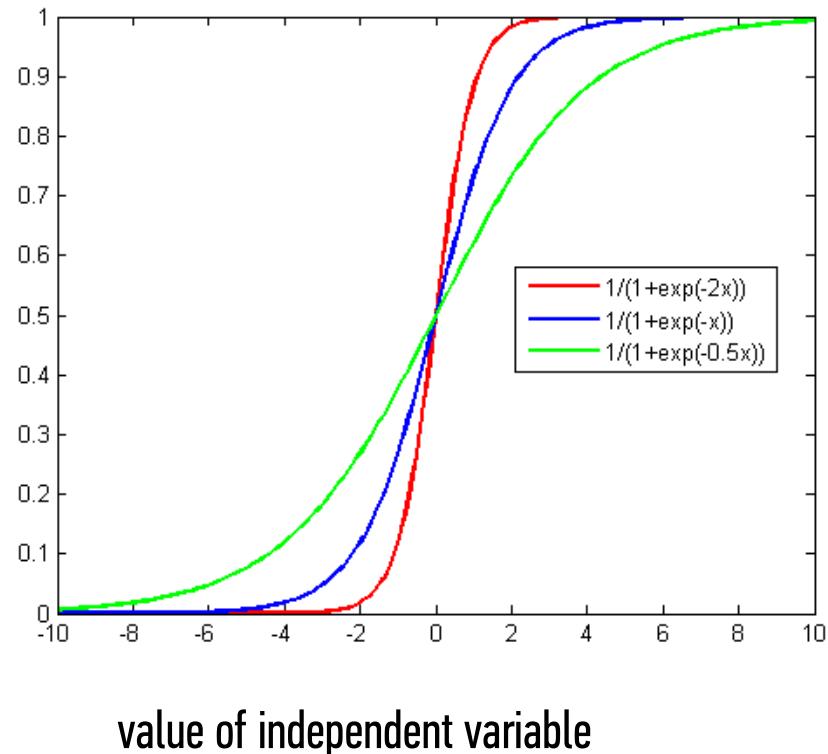
II. LOGISTIC REGRESSION

- In linear regression, we used a set of covariates to predict the value of a (continuous) outcome variable.
- In logistic regression, we use a set of covariates to predict *probabilities* of (binary) class membership.
- These probabilities are then mapped to *class labels*, thus solving the classification problem.

probability of
belonging to
class

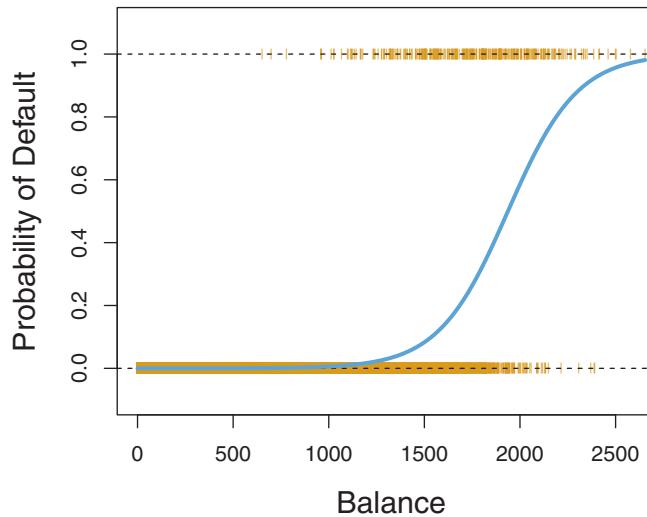
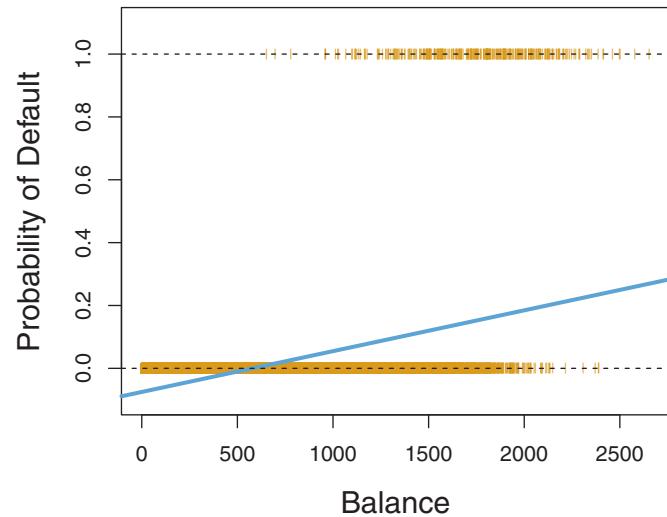


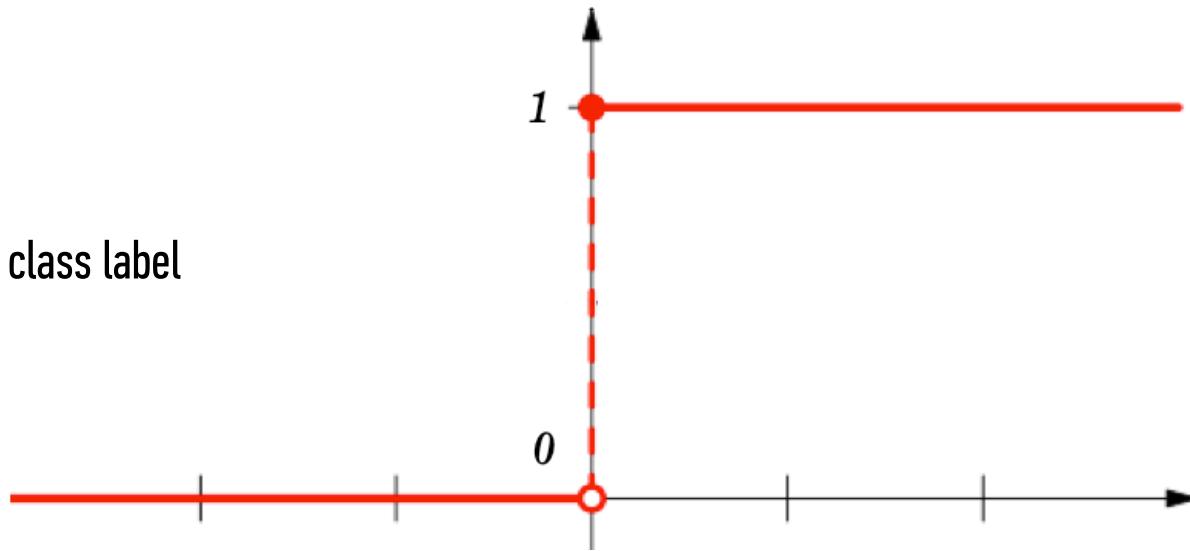
probability of
belonging to
class



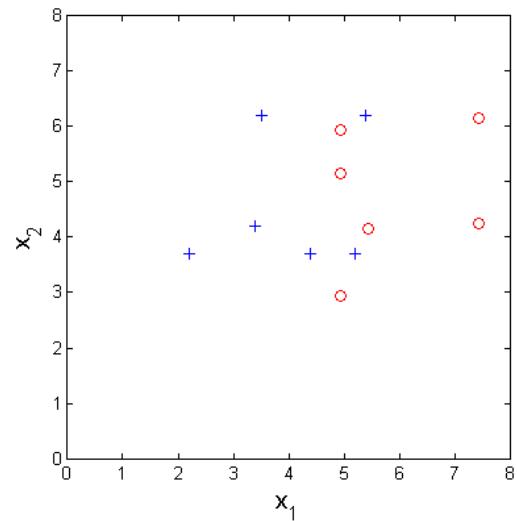
LOGISTIC REGRESSION

15

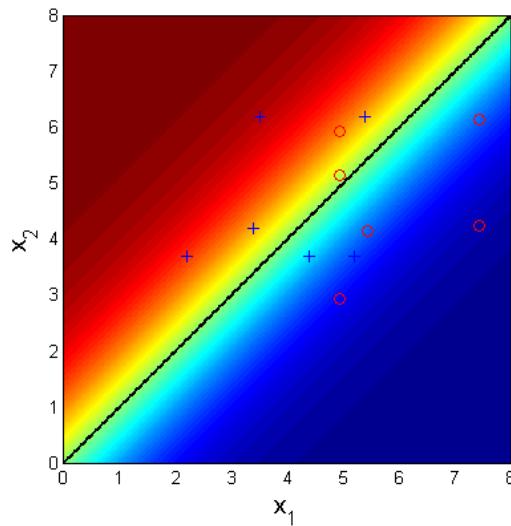




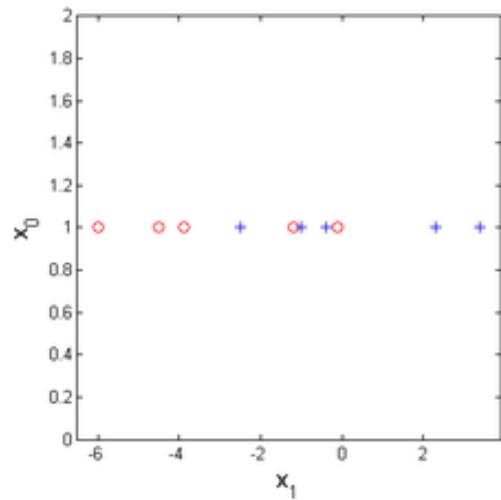
value of independent variable



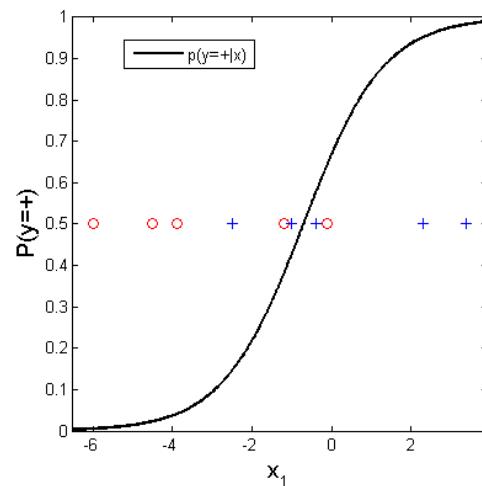
Data points



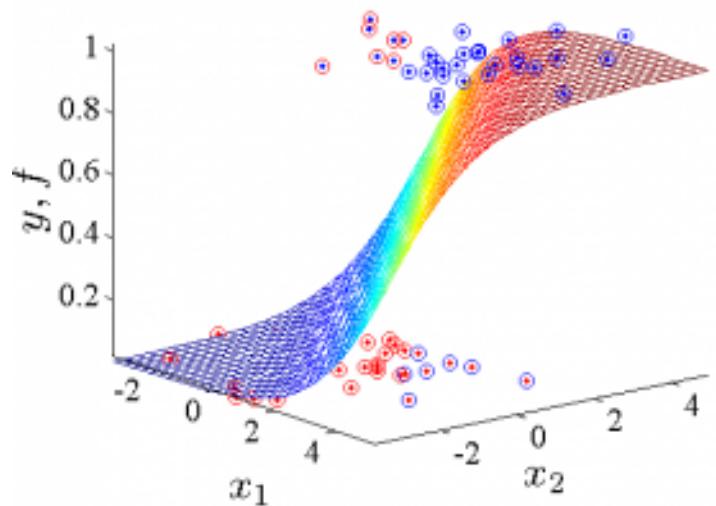
Logistic regression



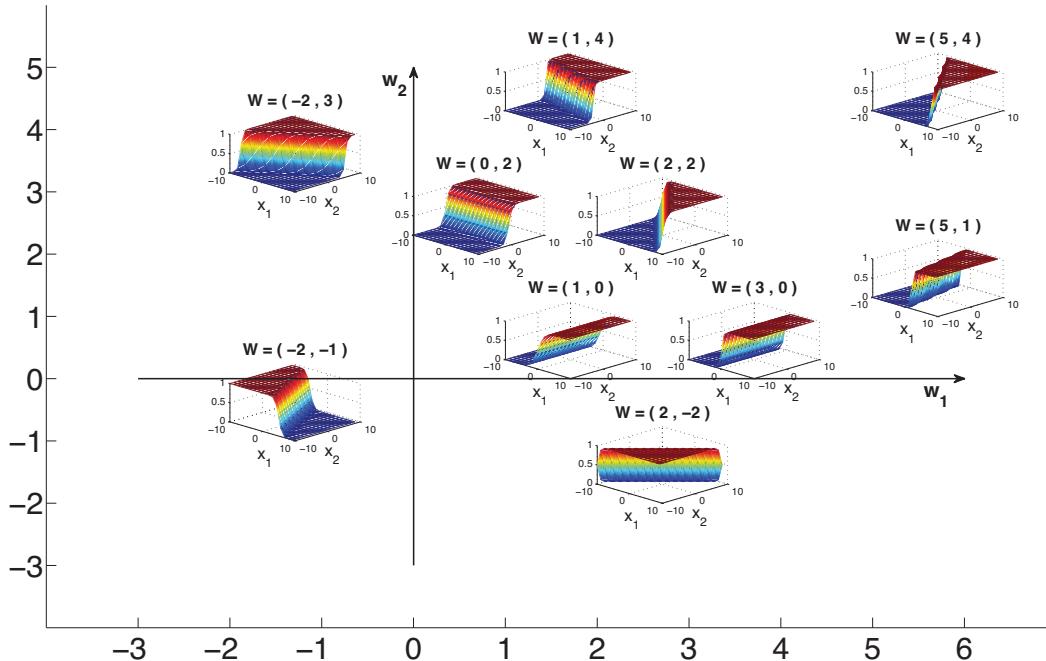
Data points



Logistic regression



Data points + logistic regression



Plots of $\text{sigm}(w_1x_1 + w_2x_2)$.

INTRO TO DATA SCIENCE

LAB: LOGISTIC REGRESSION