Midterm

10-701 Fall 2006

Outline

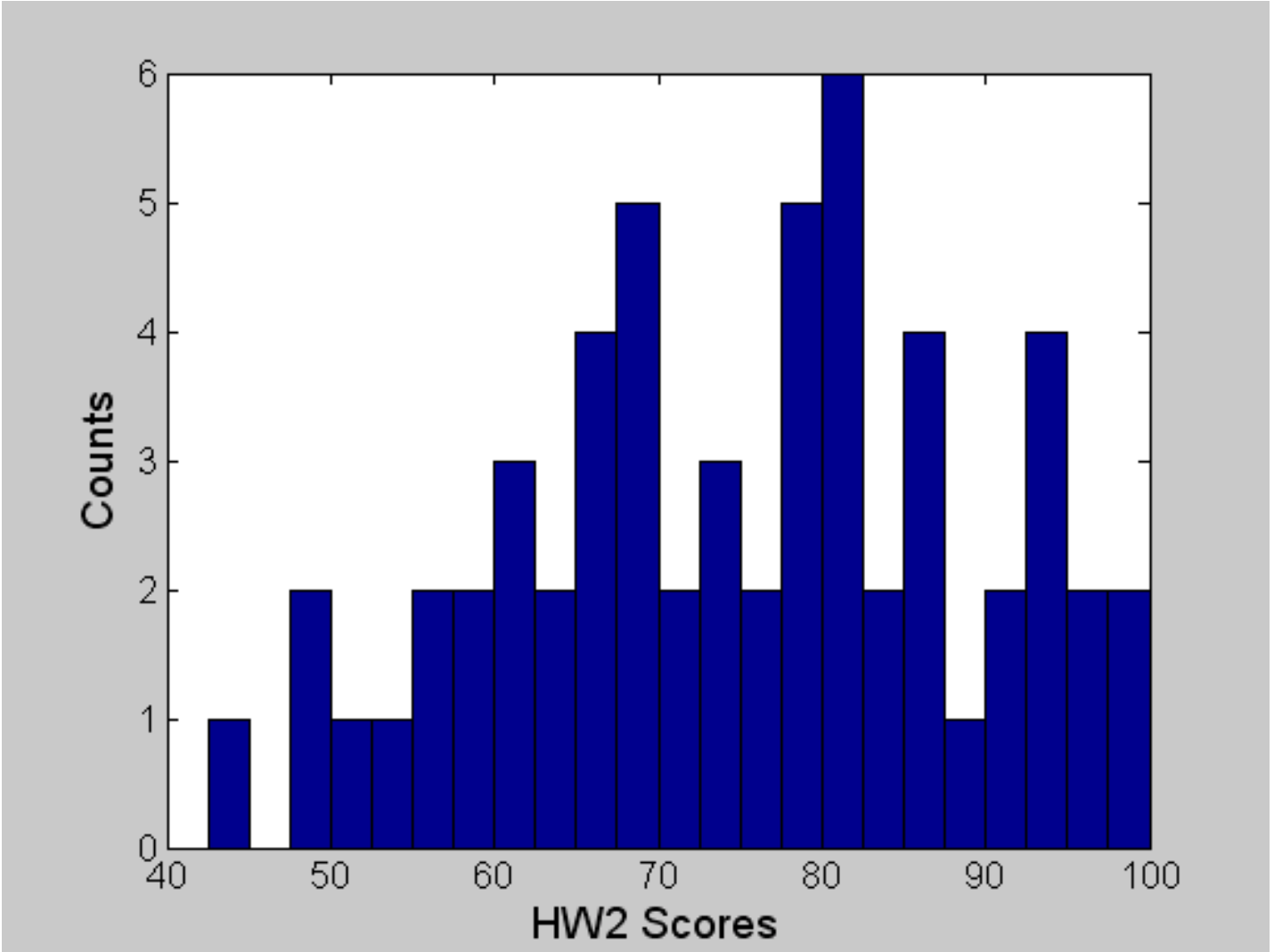
- Statistics for homework 2
- Overview of the midterm
- What you need to know
- Go over problems from previous midterms

General

- 58 homework submission including 26 late ones (last time 64 homeworks)

- Mean: 75.3 (7.9+15.5+30.6+21.2)
- Stdev: 14.0
- Median: 76.5

Histogram



Overview

- Open notes, open book
- No electronic devices (computer, cell phone, calculator)
- A lot of short problems (varying difficulty, but easier than the homeworks)
- Understand materials, look for interesting structures in problems

What you need to know

- General
- Decision trees
- Probability, MLE, MAP
- Linear regression
- Generative and discriminative classifiers (naive Bayes and logistic regression)
- Neural network
- Model selection
- Boosting
- SVM, kernel methods
- PAC learning

General

- Training error
- Test error
- Decision boundary

Decision trees

- Concept
- ID3
- Entropy, conditional entropy
- Information gain

Probability, MLE, MAP

- Axioms of probability
- Conditional probability, Bayes rule
- Independence, conditional independence
- Likelihood, MLE
- Prior, posterior, MAP

Linear regression

- Regression (vs classification)
- Estimation (gradient descent, normal equation)
- Probabilistic interpretation

Generative and discriminative classifiers

- Bayes classifier, Naive Bayes, assumptions
- Logistic regression, assumptions, regularization
- Relationship
- Generative vs discriminative classifiers

Neural networks

- What is a neural network
- What is an activation function
- Hidden layer
- Backpropagation algorithm
- Regularization

Model selection

- Purpose
- Methods: cross-validation, score
- Bias-variance decomposition
- Feature selection

Boosting

- Ensembles of classifiers
- AdaBoost

SVM

- Margin
- Support vectors
- Kernel

PAC learning

- Size of hypothesis space
- Epsilon, delta
- VC-dimension, shattering

Questions?