Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Maximum Likelihood Estimation

Introduction to Data Science Algorithms
Jordan Boyd-Graber and Michael Paul
SEPTEMBER 29, 2016

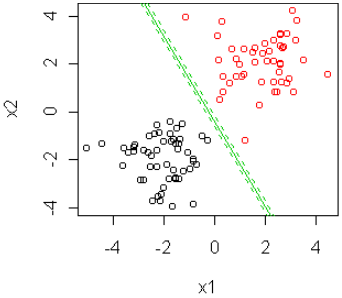
Big Pictures

- Ran through several common examples
- For existing distributions you can (and should) look up mle
- For new models, you can't (foreshadowing of later in class)

Big Pictures

- Ran through several common examples
- For existing distributions you can (and should) look up mle
- For new models, you can't (foreshadowing of later in class)
    - Classification models
    - Unsupervised models (Expectation-Maximization)
- Not always so easy

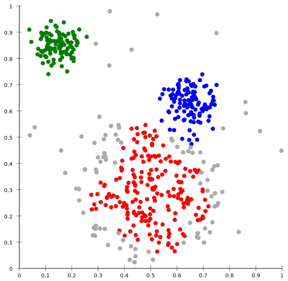
Classification



- Classification can be viewed as $p(y|x,\theta)$
- Have x, y, need θ
- Discovering θ is also problem of MLE

Clustering

- Clustering can be viewed as $p(x \mid z, \theta)$
- Have x, need z, θ
- z is guessed at iteratively (Expectation)
- θ estimated to maximize likelihood (Maximization)



Not always so easy: Bias

- An estimator is biased if $\mathbb{E}\left[\hat{\theta}\right] \neq \theta$
- We won't prove it, but the estimate for variance is biased
- Comes from estimating μ, so need to "shrink" variance

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_i (x_i - \mu)^2 \tag{1}$$

Not always so easy: Bias

- An estimator is biased if $\mathbb{E}\left[\hat{\theta}\right] \neq \theta$
- We won't prove it, but the estimate for variance is biased
- Comes from estimating μ, so need to "shrink" variance

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_i (x_i - \mu)^2 \tag{1}$$

- But we'll use biased estimate in HW2

Not always so easy: Intractable Likelihoods

- Not always possible to “solve for” optimal estimator
- Use gradient optimization (we’ll see this for logistic regression)
- Use simplier distributions as approximate (variational inference)
- Whole subfield of statistics / computer science