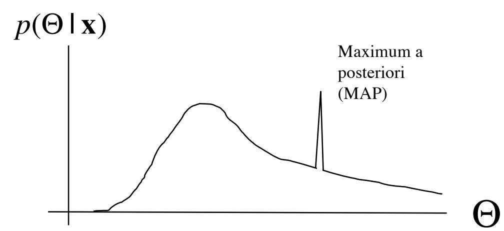
More on the Bayesian Method

- Recall that a generative probabilistic model
  - Tells a story about how stochastic data comes to be
  - Provides likelihood given data given model $p(\{\mathbf{x}_i\} | \Theta)$
  - Also provides a story for the prior $p(\Theta)$
- Bayes rule
  - Tells us how to go *from* data given model *to* model given data
  - Tell us how to combine prior knowledge and evidence from data
  - Gives a probability distribution for an answer
    - Ideal for further reasoning
    - Supports various estimates (see cartoon on next slide)
    - Supports "risk" functions

$$p(\Theta | \mathbf{x})$$

Bayesian Estimators



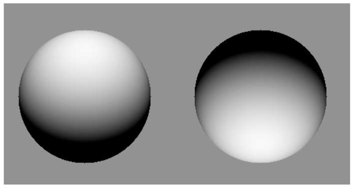
Information from Priors and Data

- Recall that vision problems do not have unique solutions!
  - We have to choose solutions suggested both by data and by what we believe (world knowledge)
  - What we believe about the world is the the prior

Simple example*

- What you know
  - John is coughing
- What do you conclude?
  - John has a cold
  - John has lung cancer
  - John has stomach problems

*Adopted from Josh Tenenbaum



Notice that the interpretation of the data is ambiguous.

The left image can be a convex with light from above, or concave with light from below.

The right image can be convex with light from below, or concave with light from above.

On average, we resolve the ambiguity by assuming that the light comes from above (prior).

Model Fitting Challenges

- Robustness
  - Squared error grows rapidly as distance increases
  - Since large distance is unlikely given Gaussian assumption, this means that either the assumption or model is likely incorrect!
- How do we know whether a point is on the line?
  - Incremental line fitting
  - K-means line fitting
  - Probabilistic with missing data

Algorithm 15.1: Incremental line fitting by walking along a curve, fitting a line to runs of pixels along the curve, and breaking the curve when the residual is too large

```
Put all points on curve list, in order along the curve
Empty the line point list
Empty the line list
Until there are too few points on the curve
  Transfer first few points on the curve to the line point list
  Fit line to line point list
  While fitted line is good enough
    Transfer the next point on the curve
      to the line point list and refit the line
  end
  Transfer last point(s) back to curve
  Refit line
  Attach line to line list
end
```

For completeness.
Not covered in 2010

Algorithm 15.2: K-means line fitting by allocating points to the closest line and then refitting.

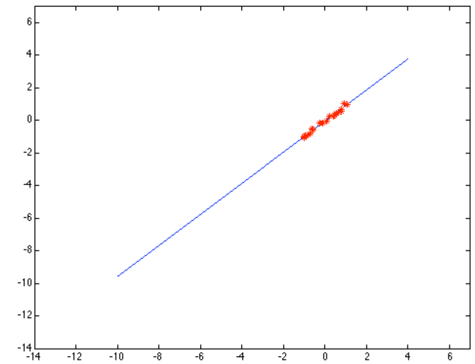
```
Hypothesize k lines (perhaps uniformly at random)
or
Hypothesize an assignment of lines to points
  and then fit lines using this assignment

Until convergence
  Allocate each point to the closest line
  Refit lines
end
```

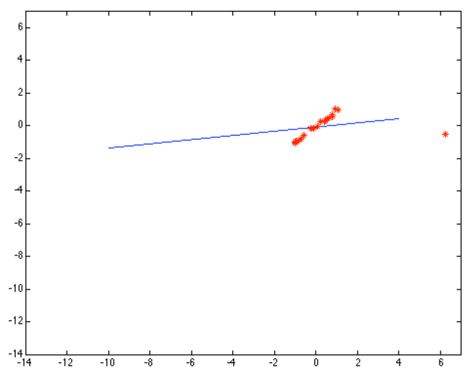
Robustness

- Squared error is a liability when model is wrong
  - One fix is EM - we'll do this shortly
  - Another is an M-estimator
    - Square nearby, threshold far away
  - A third is RANSAC
    - Search for good points

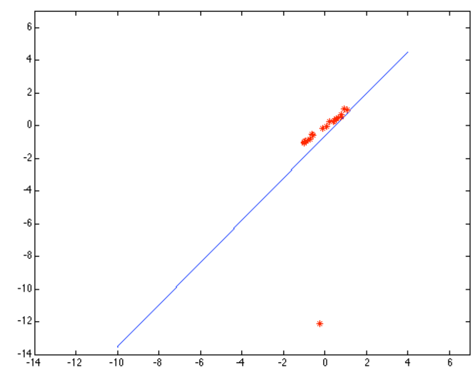
Least squares fit (good example)



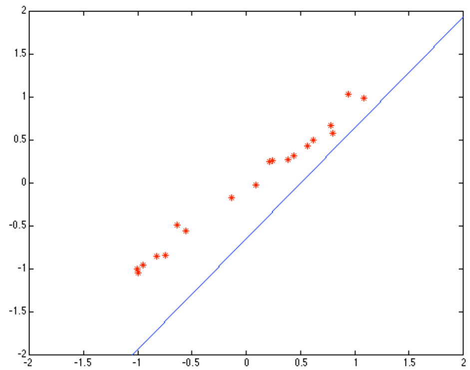
Least squares fit (destroyed by outlier)



Least squares fit (warped by outlier)



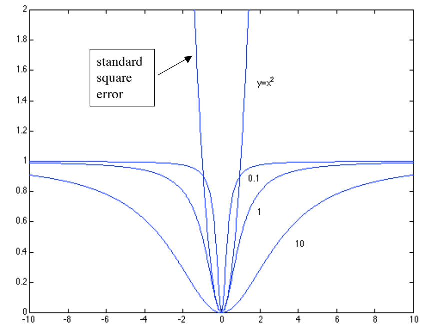
Least squares fit (previous slide zoom in)



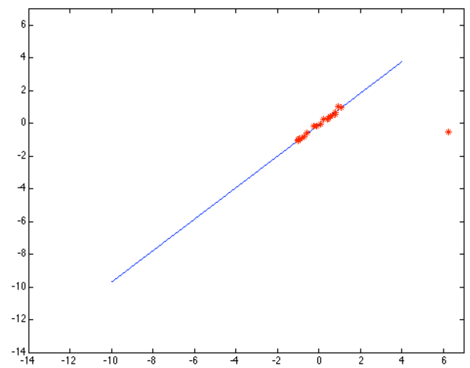
Example of a robust estimator. The effect of outliers are mitigated.
After a certain distance, errors count the same.

$y=x^2/(x^2+s^2)$

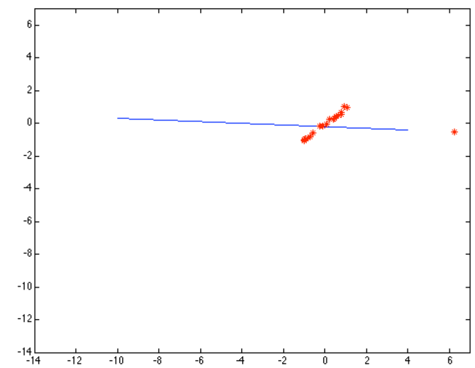
(Curve for three different values of s shown)



Line fit with estimator with good choice for s

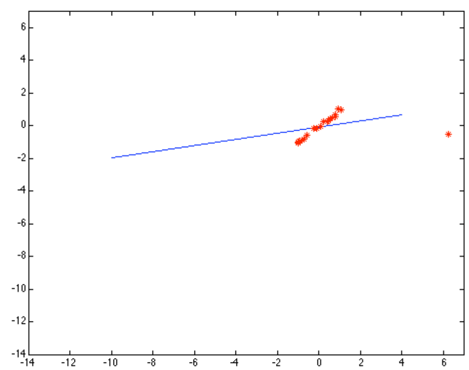


Line fit with estimator with choice for s that is too small



If s is too small, then the data is ignored too much

Line fit with estimator with choice for s that is too big



If s is too big, then we are back towards least squares

RANSAC

- Choose a minimally small subset (uniformly) at random
- Fit to that
- Anything that is close to result is signal; all others are noise
- Refit
- Measure quality
- Do this many times and choose the best

RANSAC

- How big a subset?
  - Smallest possible for the particular model (for a line, use 2 points)
- What does close mean?
  - Depends on the problem
  - Two strategies
    - Points within some fit threshold
    - Best k% points
- What is a good line?
  - One where the number of nearby points is so big it is unlikely to be all outliers (another threshold decision).

RANSAC

- How many iterations?
  - Often enough that we are likely to have a good model
    - Goes up with model complexity and belief about percentage of outliers
- Following notation from: http://en.wikipedia.org/wiki/RANSAC
  - Let w be the probability of getting an inlier.
  - Assume n points are needed for the model.
  - Suppose you want to be sure of a valid fit with probability, p.
  - Then the number of iterations, k, needed is:

$$k = \frac{\log(1-p)}{\log(1-w^n)}$$