Probabilistic Fitting

• Generative probabilistic model
  – Tells a story about how stochastic data comes to be
  – Darts fall around the center of the board, but where exactly?
  – Consider a model with parameters, \( \theta \)
  – Consider an observation, \( x_i \)
  – We denote the probability of seeing \( x_i \) under the model by:
    \[
p(x_i | \Theta)
    \]
    Read “given” or “conditioned on”
    Restricts to the case of \( \theta \)
    Defined by \( P(A | B) = \frac{P(A,B)}{P(B)} \)

Probabilistic Fitting

• Multiple observations
  – Suppose we have multiple observations, in a vector \( x \)
  – What is the probability of \( x \)?
  • If observations are independent then probability is the product of the individual observations
    – Essentially a definition, but is consistent with intuition
    – The observations are conditionally independent given the model
  • So, the probability of \( x \) is then:
    \[
p(x | \Theta) = \prod p(x_i | \Theta)
    \]

Bayes Rule

• Bayes rule:
  \[
P(A | B) = \frac{P(B | A)P(A)}{P(B)}
  \]
• Proof
  \[
P(A, B) = P(B | A)P(A) = P(A | B)P(B)
  \]
• With our notation:
  \[
P(\Theta \mid x) = \frac{P(x \mid \Theta)P(\Theta)}{P(x)}
  \]
**Probabilistic Fitting**

- If we assume uniform prior, then we can find the posterior density for the parameters by:
  \[
  p(\Theta | x) \propto p(x | \Theta)
  \]
- Now the objective is to find the parameters \( \Theta \) such that this likelihood is maximum
- Note--this is the same as finding the parameters which minimize the **negative log likelihood**

**Probabilistic fitting with independence and uniform prior**

Finding the “best” model under simple circumstances

- maximize \( p(\Theta | x) \) (one definition of best \( \Theta \))
- maximize \( p(x | \Theta) \) (by Bayes rule, uniform prior)
- minimize \( -\log(p(x | \Theta)) \) (log is monotonic increasing)
- minimize \( -\log(\prod p(x_i | \Theta)) \) (by independence)
- minimize \( -\sum \log(p(x_i | \Theta)) \) (high school math)

- Back to lines: \( ax+by+c=0 \) where \( a^2+b^2=1 \)
- Algebraic fact: Distance squared from \((x,y)\) to this line is \((ax+by+c)^2\)
- **Generative model** for lines: Choose point on line, and then, with probability proportional to \( p(d) \), normally distributed (Gaussian), go a distance \( d \) from the line.
- Now the probability of an observed \((x,y)\) is given by
  \[
  p((x,y) | \Theta) \propto \exp\left(-\frac{(ax + by + c)^2}{2\sigma^2}\right)
  \]
Lines

Convenient formula for line
\[ ax + by + c = 0 \]
where \( a^2 + b^2 = 1 \)

\[ d^2 = (ax_D + by_D + c)^2 \]

This is the generative model
It tells us \( P(\text{data} \mid \text{model}) \)

\[ p((x_D, y_D) \mid !) \propto \exp\left( -\frac{(ax_D + by_D + c)^2}{2\sigma^2} \right) \]

We have the probability density of the observed \((x,y)\) given by
\[ p((x,y) \mid \Theta) \propto \exp\left( -\frac{(ax + by + c)^2}{2\sigma^2} \right) \]

The negative log is
\[ \frac{(ax + by + c)^2}{2\sigma^2} \]

And the negative log likelihood of multiple observations is
\[ \frac{1}{2\sigma^2} \sum_i (ax_i + by_i + c)^2 \]

From the previous slide, we had that the negative log likelihood of multiple observations is given by
\[ \frac{1}{2\sigma^2} \sum_i (ax_i + by_i + c)^2 \quad (\text{where } a^2 + b^2 = 1) \]

This should be recognizable as homogeneous least squares

Thus we have shown that least squares is maximum likelihood estimation under normality (Gaussian) error statistics!

Line fitting by minimizing error == maximum likelihood estimation
Fitting curves other than lines

- In principle, an easy generalization
  - Assuming Gaussian error statistics, Euclidean distance is a good measure
  - The probability of obtaining a point, given a curve, is given by a negative exponential of distance squared
- In practice, this can be hard
  - It can be difficult to compute the distance between a point and a curve
  - Circles, ellipses, and a few others are not too hard
  - Otherwise, craft an approximation
  - §15.3 has more

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
  \[ p(\{x_i\} | \Theta) \]
- Bayes rule
  - Tells us how to go from data given model to model given data
  - Tells 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
  \[ p(\Theta | x) \]

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 prior