Least squares method: linear regression

Ramani Duraiswami,
Dept. of Computer Science
Fitting data to a model

• Practical science involves lots of fitting of data to models

• Difference between fitting and interpolation?
  – Interpolation, the fit function passes through the point
  – Fitting, the fit function satisfies some error criterion

• Tasks arise commonly in science
  – Fit straight lines and curves to data
  – More generally fit data to a parametric model

• Parametric: Model contains parameters
  – Job of fitting is to estimate the parameters that “best” make the model fit the data
  – “best” → define best

• Simplest example of model fitting problem
  – Linear regression
Models

- Have a certain model structure
  - E.g., “linear” “quadratic” “trigonometric” “Gaussian”
- Models have specifiable parameters

<table>
<thead>
<tr>
<th>Model Structure</th>
<th>Data $$(x_i,y_i)$$</th>
<th>Param. $$a,b,c$$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Straight line:</td>
<td>$a x + by +c =0$</td>
<td>$a,b,c$</td>
</tr>
<tr>
<td>Polynomial:</td>
<td>$y=c_0+c_1x+...+c_nx^n$</td>
<td>$c_0, c_1, ..., c_n$</td>
</tr>
</tbody>
</table>
- General model
- $y(x) \simeq \beta_1 \phi_1(x) + ... + \beta_n \phi_n(x) +$ 
- E.g., $\simeq \beta_1 x + ... + \beta_n \phi_n(x) +$ 
- Data $$(y_i,x_i)$$ and 
- $y=\Phi\beta$ 
- Solve $\beta=\Phi\backslash y$
Relationships among Variables

• In much science we seek relations between variables

One variable is used to “explain” another variable

X Variable
  Independent Variable
  Explaining Variable
  Exogenous Variable
  Predictor Variable

Y Variable
  Dependent Variable
  Response Variable
  Endogenous Variable
  Criterion Variable
Simple Least-Squares Regression

If we had errorless predictions:

\[ Y = a + bX \]
Estimated Regression Line

Equation of the Regression Line:
\[ \hat{Y} = a + bX \]

\[ e_i = y_i - \hat{y}_i \]
Estimated Regression Line

$Y$

$X$

errors/residuals
Linear Systems

Square system:
• unique solution
• Gaussian elimination

Rectangular system ??
• underconstrained: infinity of solutions
• overconstrained: no solution

Minimize \(|Ax-b|^2\)

For straight line fitting we need to find two parameters. However each of the data points provides an estimate of the error.
How do we find $a$ and $b$?

In Least-Squares Regression:

Find $a, b$ to minimize the sum of squared errors/residuals

$$
\sum_{i=1}^{N} (e_i)^2 = \sum_{i=1}^{N} (y_i - [bx_i + a])^2
$$
Least Squares for more complex models

- Number of equations and unknowns may not match
- Look for solution by minimizing some cost function
- Simplest and most intuitive cost function: $\|Ax - b\|_2$
- Define for each data point $x_i$ a residual $r_i$
- Minimize $\sum_i r_i$ with respect to $x_i$

$$\sum_i r_i = \sum_j (A_{ij}x_j - b_i) = \sum_k (A_{ik}x_k - b_i)$$

\[
\frac{\partial}{\partial x_i} (A_{ij}x_j - b_i) \cdot (A_{ik}x_k - b_i) = 0
\]

\[
(A_{ij} \delta_{jl}) \cdot (A_{ik}x_k - b_i) + (A_{ij}x_j - b_i) \cdot (A_{ik} \delta_{kl}) = 0
\]

\[
A_{il} \cdot (A_{ik}x_k - b_i) + (A_{ij}x_j - b_i) \cdot A_{il} = 2 (A_{il}A_{ik}x_k - A_{il}b_i) = 0
\]

\[
A_{il} A_{ik} x_k = A_{il} b_i
\]
Other Norms

- Here we fit using the “least-squares” or $L_2$ norm
- Could minimize the residual in other norms
- For example we may have more confidence in some data, and want to be sure that their residual is lower
  - Attach a weight to each residual
    \[ \| r \|^2_w = \sum_{i=1}^{m} w_i r_i^2 \]
- Or we may like the $1$-norm or infinity norm better
  \[ \| r \|_1 = \sum_{i=1}^{m} |r_i| \]
  \[ \| r \|_\infty = \max_i |r_i| \]
Normal equations

• The system $A^t A x = A^t b$
  is called the Normal equations

• Can solve least squares problems using these

• For $A$ size $m \times n$ and $x$ of size $n$ and $b$ of size $m$ what are
  the dimensions of the normal equations?
  – $n \times n$

• Solve via LU decomposition

• Is this a good idea?
  – Somewhat expensive as we have to form $A^t A$ which involves
    matrix multiplication and then solution
  – More importantly it is poorly conditioned
  – $\text{cond}(A^t A) = (\text{cond}(A))^2$