Lecture 19

Linear Models for Regression

Simple example of fitting data to a straight line order $n=1$

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \text{, or } V_{(3\times2)}\, p_{(2\times1)} = y_{(3\times1)}$$

Given training data samples (x, y): (2, 5), (3, 7), (4, 9), the system of equations (with 2 unknowns and 3 equations):

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} p_1 \\ p_2 \end{bmatrix}}_{p} = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix}$$

$p(x) = p_1 x + p_2$

if $n=1$ $\Rightarrow \begin{cases} 2p_1 + p_2 = 5 \\ 3p_1 + p_2 = 7 \\ 4p_1 + p_2 = 9 \end{cases}$

- p = polyfit (x,y,n) returns the coefficients for a polynomial p(x) of degree n that is a best fit (in a least-squares sense) for the data in y. The coefficients in p are in descending powers, and the length of p is n+1.

$$p(x) = p_1 x^n + p_2 x^{n-1} + \ldots + p_n x + p_{n+1}.$$

- polyfit uses x to form a Vandermonde matrix V with m = length(x) rows and (n+1) columns, resulting in the linear system below, which polyfit solves with p = V\y = pinv(V) * y.

$$\underbrace{\begin{pmatrix} x_1^n & x_1^{n-1} & \cdots & 1 \\ x_2^n & x_2^{n-1} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ x_m^n & x_m^{n-1} & \cdots & 1 \end{pmatrix}}_{m \times (n+1)} \underbrace{\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n+1} \end{pmatrix}}_{(n+1) \times 1} = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}}_{m \times 1},$$

$n=0$

Goal: Find a solution vector p such that the approximation error (squared) below is minimized: $E^2(p) = \|Vp - y\|^2$.

$(2p_1 + p_2 - 5)^2$

$= 4p_1^2 + p_2^2 + 25 + 4p_1p_2 - 20p_1 - 10p_2$

Symbolic Matrix Operations

$$E^2(p) = \underbrace{\|Vp - y\|^2}_{1\times 1} = \underbrace{(Vp-y)^T}_{(1\times m)}\underbrace{(Vp-y)}_{(m\times 1)} = (p^T V^T - y^T)(Vp - y)$$

$$= p^T V^T V p - p^T V^T \mathrm{y} - \mathrm{y}^T V p + \mathrm{y}^T \mathrm{y}$$

$(Vp)^T = p^T V^T$

```
>> syms p1 p2 p E2(p1,p2)
p =[p1;p2];
V = [2, 1; 3, 1; 4, 1];
y = [5 7 9]';
E2(p1,p2) = (p.')*(V')*V*p -(p.')*(V')*y -
y'*V*p +y'*y;

>> simplify(E2)
ans = 29*p1^2 + 18*p1*p2 - 134*p1 +
3*p2^2 - 42*p2 + 155
```

$$f(x) = (Vx-y)^T(Vx-y) = \|Vx-y\|^2$$

$$= 29p_1^2 + 18p_1p_2 - 134p_1 - 42p_2 + 155 + 3p_2^2$$

$$\frac{\partial f}{\partial p_1} = 58p_1 + 18p_2 - 134 = 0, \quad \frac{\partial f}{\partial p_2} = \cdots = 0$$

$\nabla E^2(p) = \begin{bmatrix} \frac{\partial E^2}{\partial p_1} \\ \frac{\partial E^2}{\partial p_2} \end{bmatrix} = \nabla_p(p^T V^T V \mathrm{p} - p^T V^T \mathrm{y} - \mathrm{y}^T V p + \mathrm{y}^T \mathrm{y}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, in order to determine the *critical point* that can potentially minimize $E^2(p)$, where

$$\nabla_p(p^T V^T V p) = 2(V^T V)p, \quad \nabla_p(p^T V^T \mathrm{y}) = \nabla_p(y^T V p) = V^T \mathrm{y}, \quad \nabla_p(y^T y) = 0$$

Thus $(V^T V)p - V^T \mathrm{y} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, or $V^T V p = V^T \mathrm{y}$ (Normal Equation in Statistics)

For a scalar α given by a quadratic form: $\alpha = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$ ← $p^T (V^T V) p$, $A = V^T V$, p^T ↑ x^T, p ↑ x

For the special case $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$, then $\underbrace{\frac{\partial[\mathbf{x}^{\mathrm{T}}(\mathbf{A}+\mathbf{A}^{\mathrm{T}})]}{\partial \mathbf{x}}}_{\text{column vector}} = (2\mathbf{x}^{\mathrm{T}}\mathbf{A})^T = 2A^T x$

$$(V^T V)^T = V^T (V^T)^T = V^T V$$

And.

$$\overbrace{p^T V^T y}^{1\times 1}, \quad p^T: 1\times(n+1), \quad V^T: (n+1)\times m, \quad y: m\times 1$$

$$\underbrace{(p^T V^T y)^T}_{1\times 1} = \underbrace{y^T V p}_{1\times 1} \Rightarrow \nabla_p (p^T V^T y) = \overbrace{V^T y}^{(n+1)\times 1\text{ : column vector}}, \quad V^T: (n+1)\times m, \quad y: m\times 1$$

Normal Equation:

Thus $(V^TV)p - V^Ty = \begin{bmatrix}0\\0\end{bmatrix}$, or $\underset{(n+1)\times m}{V^T}\,\underset{m\times(n+1)}{V}\,\underset{(n+1)\times 1}{p} = V^Ty$ (Normal Equation in Statistics)

V^TV is invertible when the columns of V are linearly independent.
Best estimate (in least square sense): $\hat{p} = [(V^TV)^{-1}V^T]\,y = \text{pinv}(V)\,y$

```
>> V = [2, 1; 3, 1; 4, 1];
>> y = [5 7 9]';
```

```
>> inv(V'*V)*V'*y
>> pinv(V)*y
```

```
ans =
  2.0000
  1.0000
```

$p = \begin{pmatrix}2\\1\end{pmatrix}$

$$V^TVp = V^Ty$$

$$\underbrace{(V^TV)^{-1}(V^TV)}_{I}\,p = (V^TV)^{-1}V^Ty$$

$$p = \left[(V^TV)^{-1}V^T\right]y$$

```
V =

     2     1
     3     1
     4     1
```

```
y =

     5
     7
     9
```

Hessian Matrix (Derivative of Gradient)

$$\nabla_p^2 E^2(p) = \begin{bmatrix} \dfrac{\partial^2 E^2}{\partial p_1^2} & \dfrac{\partial^2 E^2}{\partial p_1 \partial p_2} \\ \dfrac{\partial^2 E^2}{\partial p_2 \partial p_1} & \dfrac{\partial^2 E^2}{\partial p_2^2} \end{bmatrix} = \nabla_p \begin{bmatrix} \dfrac{\partial E^2}{\partial p_1} \\ \dfrac{\partial E^2}{\partial p_2} \end{bmatrix}^T = \nabla_p (2V^TVp - 2V^Ty)^T$$

$$= \nabla_p (2p^TV^TV - 2y^TV) = 2V^TV$$

- V^TV is always symmetric and positive definite (with all eigenvalues being positive, all pivots being positive), thus $\hat{p} = [(V^TV)^{-1}V^T]\,y$ is not only a critical point, but also a *local* minima.
- In addition, due to the Hessian being a (everywhere in general) positive definite matrix, $E^2(p)$ is a convex function, and $\hat{p}$ is also a **global minima**.

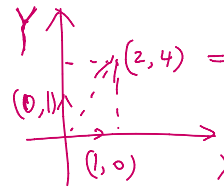
Geometric Interpretation

$$\begin{array}{cc} x_1 & x_2 \end{array}$$
$$\begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix} = y, \text{ Solution: } \hat{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$y = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix} = p_1 \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} + p_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix}\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = y$$

$$x_1 p_1 + x_2 p_2 = y$$



$$(2,4) = 2\begin{pmatrix} 1 \\ 0 \end{pmatrix} + 4\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$(0,1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0$$

In this specific example (with zero estimation error), the 3×1 vector y happens to be in the **column space** of the matrix V, with the solution 2×1 vector $\hat{p}$ containing the components (linear combination coefficients).

- Consider the following (slightly changed) least square problem:

V: $m \times (n+1)$, V^T: $(n+1) \times m$, V^TV: $(n+1) \times (n+1)$

$$\begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}\begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 9 \end{bmatrix} = y, \text{ then } V^TV = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 29 & 9 \\ 9 & 3 \end{bmatrix}$$

$$\hat{p} = (V^TV)^{-1}V^Ty = \begin{bmatrix} 29 & 9 \\ 9 & 3 \end{bmatrix}^{-1}\begin{bmatrix} 2 & 3 & 4 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 5 \\ 6 \\ 9 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{11}{6} & \frac{1}{3} & -\frac{7}{6} \end{bmatrix}\begin{bmatrix} 5 \\ 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 2 \\ \frac{2}{3} \end{bmatrix}$$

$$y = \begin{bmatrix} 5 \\ \cancel{7}\,6 \\ 9 \end{bmatrix} \approx \begin{bmatrix} 4\frac{2}{3} \\ 6\frac{2}{3} \\ 8\frac{2}{3} \end{bmatrix} = p_1 \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} + p_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 2\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} + \frac{2}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$