INTRODUCTORY APPLIED MACHINE LEARNING

Yan-Fu Kuo
Dept. of Biomechatronics Engineering
National Taiwan University

Today:

- Math review

Vector Space

- Vector in R^n is an ordered set of n real numbers
- Column vector: $\boldsymbol{x} = [x_1, \quad \dots, \quad x_n]^{\mathrm{T}} \in R^n$
- Example: $\boldsymbol{x} = [1, 2, 3, 4]^{\mathrm{T}},\ n = 4$
- $\boldsymbol{x}, \boldsymbol{y} \in R^n \rightarrow\ \boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x} \in R^n$
- $\boldsymbol{x} \in R^n, \alpha \in R \rightarrow \alpha\boldsymbol{x} \in R^n$
- $\boldsymbol{x} \in R^n,\ \alpha, \beta \in R \rightarrow (\alpha + \beta)\boldsymbol{x} = \alpha\boldsymbol{x} + \beta\boldsymbol{x} \in R^n$

Vector Norms

- A norm of a vector $\|\boldsymbol{x}\|_p$ is informally a measure of the "size" of the vector

$$\|\boldsymbol{x}\|_p \equiv \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}$$

- Common norms:
  - L_1 norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
  - Euclidean (L_2) norm: $\|\boldsymbol{x}\|_2 = (\sum_{i=1}^{n} {x_i}^2)^{1/2} = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$
  - Infinite (L_∞) norm: $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$

- Cauchy-Schwartz inequality: $\boldsymbol{x}, \boldsymbol{y} \in R^n$, $|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}| \leq \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2$

Vector Product

- $\boldsymbol{x}, \boldsymbol{y} \in R^n$
- Inner product:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i \in R$$

- Outer product:

$$\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 & \dots & x_1 y_n \\ \vdots & \ddots & \vdots \\ x_n y_1 & \cdots & x_n y_n \end{bmatrix} \in R^{n \times n}$$

Matrix

- An m-by-n matrix is an object in $R^{m\times n}$ with m rows and n columns, each entry filled with a real number
- $\boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \in R^{m\times n}$
- $\alpha \in R \rightarrow \alpha\boldsymbol{X} = \begin{bmatrix} \alpha x_{11} & \cdots & \alpha x_{1n} \\ \vdots & \ddots & \vdots \\ \alpha x_{m1} & \cdots & \alpha x_{mn} \end{bmatrix} \in R^{m\times n}$
- $\boldsymbol{X} \in R^{m\times n}, \boldsymbol{Y} \in R^{n\times q} \rightarrow \boldsymbol{XY} \in R^{m\times q}$

Matrix Transpose

- $\boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \rightarrow \boldsymbol{X}^{\mathrm{T}} = \begin{bmatrix} x_{11} & \cdots & x_{m1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{mn} \end{bmatrix} \in R^{n \times m}$
- $(\boldsymbol{X}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{X}$
- $\boldsymbol{Y} \in R^{m \times n} \rightarrow (\boldsymbol{X} + \boldsymbol{Y})^{\mathrm{T}} = \boldsymbol{X}^{\mathrm{T}} + \boldsymbol{Y}^{\mathrm{T}}$
- $(\alpha \boldsymbol{X})^{\mathrm{T}} = \alpha(\boldsymbol{X}^{\mathrm{T}})$
- $(\boldsymbol{XYZ})^{\mathrm{T}} = \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}$

Inverse of Matrix

- $\boldsymbol{X} \in R^{n \times n}$
- $\exists \boldsymbol{Y} \in R^{n \times n}$ s.t. $\boldsymbol{XY} = \boldsymbol{YX} = \boldsymbol{I}_n$
  - If $\boldsymbol{X}$ is invertible or nonsingular, $\boldsymbol{Y}$ is the inverse of $\boldsymbol{X}$
- $\boldsymbol{A}, \boldsymbol{B} \in R^{n \times n}$ nonsingular, then $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
- $\boldsymbol{Z} \in R^{n \times n}$, suppose $\boldsymbol{XY} = \boldsymbol{ZY}$, $\boldsymbol{Y} \neq \boldsymbol{0}$
  - If $\boldsymbol{Y}$ is invertible, then $\boldsymbol{X} = \boldsymbol{Z}$
  - If $\boldsymbol{Y}$ is not invertible, then $\boldsymbol{X} \neq \boldsymbol{Z}$
- Example: $\boldsymbol{X} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$, $\boldsymbol{Z} = \begin{bmatrix} 2 & 4 \\ 2 & 3 \end{bmatrix}$, $\boldsymbol{Y} = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix}$

$$\boldsymbol{XY} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} -2 & 4 \\ -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} = \boldsymbol{ZY}$$

Vector and Matrix Derivate

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}$$

$$\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{b}}{\partial \boldsymbol{X}} = \boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$$

$$\frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{a}}{\partial \boldsymbol{X}} = \frac{\partial \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a}}{\partial \boldsymbol{X}} = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$$

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \left(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}\right)\boldsymbol{x}$$

- More: https://en.wikipedia.org/wiki/Matrix_calculus

Function Derivative

- Let $\boldsymbol{x} = [x_1 \cdots x_n]^{\mathrm{T}}$, if $f: \Re^n \longrightarrow \Re$ differentiable, then

$$\nabla f = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

- First-order necessary condition:

  If x^* a <u>local minimizer</u> of f, then $\nabla f(x*) = 0$

Gradient Example MATLAB Code

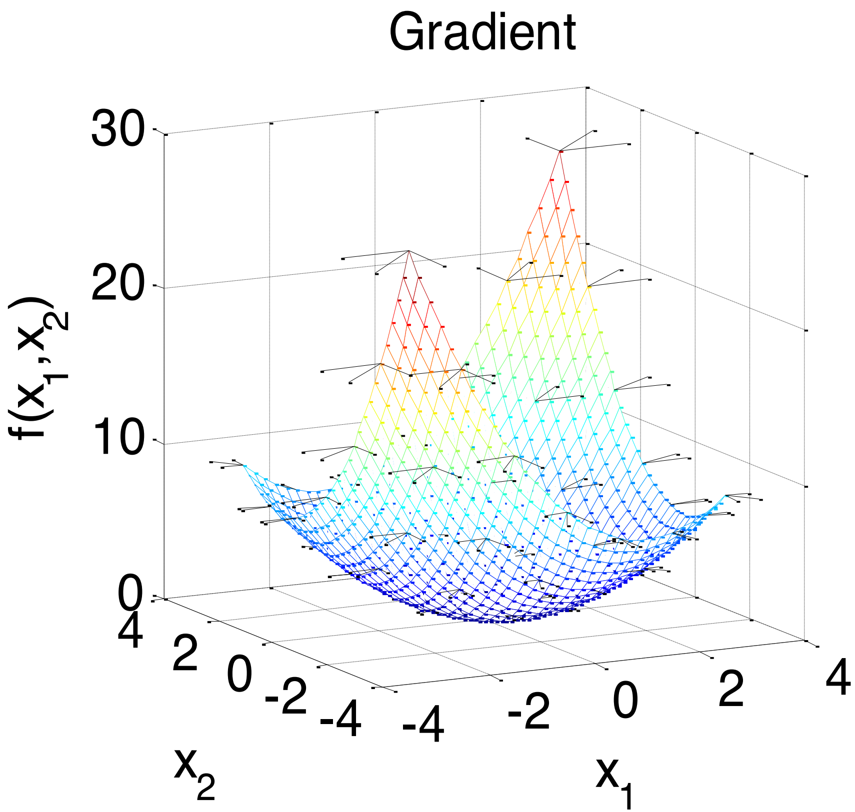
```
% generate gradient vectors
[x, y] = meshgrid(-3:1:3,-3:1:3);  z = x.^2 + x.*y + y.^2;
x_vert = x(:, 1);  y_vert = y(:, 1);  z_vert = z(:, 1);
for i=2:length(x)
    x_vert = vertcat( x_vert, x( :, i));  y_vert = vertcat( y_vert, y( :, i));
    z_vert = vertcat( z_vert, z( :, i));
end

for i=1:length(x_vert)
    gradx_vert(i) = x_vert(i)*2 + y_vert(i);  grady_vert(i) = x_vert(i) +
y_vert(i)*2;
end

gradx_vert = gradx_vert'; grady_vert = grady_vert';
% plot mesh and gradient
[x, y] = meshgrid(-3:.2:3,-3:.2:3);  z = x.^2 + x.*y + y.^2; mesh( x, y, z);
xlim([-4 4]); xlabel('x_1', 'FontSize', 16);  ylim([-4 4]);
ylabel('x_2', 'FontSize', 16);  set( gcf, 'Color', 'w');
zlabel('f(x_1,x_2)', 'FontSize', 16); title('Gradient');  set(gca,'FontSize', 16);

hold on;
arrow3( [x_vert y_vert z_vert], [x_vert+gradx_vert/7 y_vert z_vert], [], .4, 1.5);
arrow3( [x_vert y_vert z_vert], [x_vert y_vert+grady_vert/7 z_vert], [], .4, 1.5);
arrow3( [x_vert y_vert z_vert], [x_vert+gradx_vert/7 ...
    y_vert+grady_vert/7 z_vert], [], .4, 1.5);
```

Example Figure



Reference

- Linear Algebra: Determinants, Inverses, Rank
- K. Petersen and M. Pedersen, The Matrix Cookbook