Introduction to Statistical Machine Learning

STAT 5630

Minh Pham

[1]Department of Statistics
University of Virginia

Jan 20th, 2016

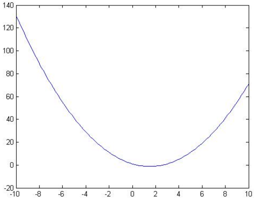
Outline

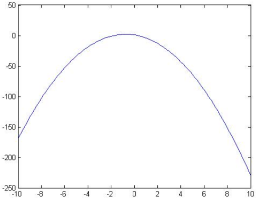
Some Matlab syntax

- For loops and While loops
- Conditional: if
- Creating functions

Quadratic function

- Quadratic function: is one of the form $f(x) = ax^2 + bx + c$, where a, b, and c are numbers with a not equal to zero.
  E.g: $x^2 - 3x + 1.$, $-2x^2 - 3x + 1$





- Matrix form: Given $Q \in \mathbb{R}^{n \times n}$, $b, x \in \mathbb{R}^{n \times 1}$:
  $f(x) = \frac{1}{2}x^T Q x + b^T x$

Quadratic function

- Convince yourself that: $x^T Qx = \sum_{i=1}^n \sum_{j=1}^n Q_{ij}x_ix_j$
- $x^T Qx = x^T(\frac{1}{2}Q + \frac{1}{2}Q^T)x$.
- Eg: Least square loss : $f(\beta) = \frac{1}{2}\|y - X\beta\|_2^2$
- Eg: norm function: $\|x\|^2 = x^T x$.

Positive semidefinite matrices

- A square symmetric maxtrix Q is positive definite (PD) if for all non zero $x \in \mathbb{R}^n, x^T Qx > 0$. The set of all PD matrices is usually denoted as : $\mathbb{S}^n_{++}$.
- A square symmetric maxtrix Q is positive semidefinite (PSD) if for all non zero $x \in \mathbb{R}^n, x^T Qx \geq 0$. The set of all PD matrices is usually denoted as : $\mathbb{S}^n_{+}$.
- A square symmetric maxtrix Q is negative definite (ND) if for all non zero $x \in \mathbb{R}^n, x^T Qx < 0$.
- A square symmetric maxtrix Q is negative semidefinite (NSD) if for all non zero $x \in \mathbb{R}^n, x^T Qx \leq 0$.
- A PD or ND matrix is always full rank (invertible).

Eigen values and eigen vectors

- Given a square matrix $Q \in \mathbb{R}^{n\times n}$, λ is an eigenvalue of Q and $x \in \mathbb{R}^n$ is a corresponding eigenvector of Q if : $Qx = \lambda x$.
- For any scalar c: $Q(cx) = \lambda(cx)$.
- $(\lambda\mathbb{I} - A)x = 0, x \neq 0$. or $det(\lambda\mathbb{I} - A)) = 0$.
- Rank of A is equal to the number of non-zero eigenvalues of Q.
- Trace of a matrix: $Tr(Q) = \sum_{i=1}^{n} Q_{ii} = \sum_{i=1}^{n} \lambda_i$.
- Q has at most n eigen values. Stack eigenvectors to $X \in \mathbb{R}^{n\times n}$ and let Λ be a diagonal matrix with diagonal entries are eigenvalues of Q: $QX = X\Lambda$.

Eigenvalues and eigenvectors of symmetric matrices

- Eigenvectors of symmetric matrices are orthogonal:
  $Q\lambda_1 = \lambda_1 x_1$,
  $Q\lambda_2 = \lambda_2 x_2, x_2^T \lambda_1 x_1 = x_2^T (Qx_1) = \lambda_2 x_1^T x_2 \rightarrow x_1^T x_2 = 0$
- Eigenvalues of symmetric matrices are real.
- Positive definite matrices if and only if all eigenvalues $>$ 0.

Matrix calculus

- $f : \mathbb{R}^{m\times n} \to R$ is a function that takes an input as a matrix A of size $m \times n$ and output a real number. Gradient of f is the matrix of partial derivatives that has size $m \times n$: $\nabla f(A) = \begin{pmatrix} \frac{\partial f(A)}{\partial A_{11}} & \frac{\partial f(A)}{\partial A_{12}} & \dots & \frac{\partial f(A)}{\partial A_{1n}} \\ \frac{\partial f(A)}{\partial A_{21}} & \frac{\partial f(A)}{\partial A_{22}} & \dots & \frac{\partial f(A)}{\partial A_{2n}} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f(A)}{\partial A_{n1}} & \frac{\partial f(A)}{\partial A_{n2}} & \dots & \frac{\partial f(A)}{\partial A_{nn}} \end{pmatrix}$.
  So $\nabla f(A) = \frac{\partial f(A)}{\partial A_{ij}}$.
- $\nabla(f+g)(x) = \nabla f(x) + \nabla g(x)$, $\nabla(af(x)) = a\nabla f(x)$

Matrix calculus - Hessian matrix

- Given $f : \mathbb{R}^n \to R$, the Hessian matrix, $\nabla^2 f(x)$ (H) is the $n \times n$ matrix of partial derivatives:

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial f(A)}{\partial x_1 x_1} & \frac{\partial f(A)}{\partial x_1 x_2} & \dots & \frac{\partial f(A)}{\partial x_1 x_n} \\ \frac{\partial f(A)}{\partial x_2 x_1} & \frac{\partial f(A)}{\partial x_2 x_2} & \dots & \frac{\partial f(A)}{\partial x_2 x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f(A)}{\partial x_n x_1} & \frac{\partial f(A)}{\partial x_n x_2} & \dots & \frac{\partial f(A)}{\partial x_n x_n} \end{pmatrix}.$$

Gradient and Hessian of simple functions

- Let $f(x) = b^T x = \langle b, x \rangle = dot(b, x)$, then $\nabla f(x) = b$.
- Let $f(x) = \frac{1}{2} x^T Q x + b^T x$, then $\nabla f(x) = (Q + Q^T)x + b$.