CS5014 Machine Learning

$x^T Ax$ Quadratic form

Lei Fang

01/02/2021

1 Introduction

A quadratic form is defined as

$$x^T Ax,$$

where $x = \begin{bmatrix} x_1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}$ is a $n \times 1$ (column) vector, A is a $n \times n$ square matrix (the corresponding i-th row and j-th column entry is a_{ij}. Therefore, $x^T Ax$ is scalar function of input x. According to matrix multiplication rule, the quadratic form can be expanded as

$$x^T Ax = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

essentially all the second order products between $x_i x_j$ and weighed by a_{ij}.

Example 1. $f(x) = x_1^2 + x_2^2$ is a quadratic form, as

$$f(x) = x^T \begin{bmatrix} 1, 0 \\ 0, 1 \end{bmatrix} x = x^T x$$

. Its surface plot in R^3 and contour plot is show below in Fig 1. A contour plot shows all the levels sets: *i.e.* all the $x \in R^n$ in the input space such at $f(x) = c$ is a constant.

Example 2. $f(x) = 4x_1^2 + x_2^2$ is a quadratic form as well:

$$f(x) = [x_1, x_2] \begin{bmatrix} 4, 0 \\ 0, 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

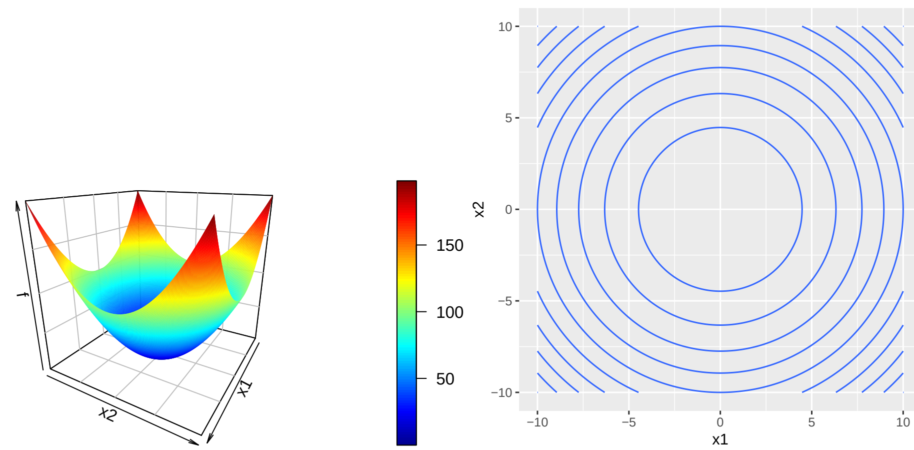

Figure 1: surface plot of a circular paraboloid

The plots shown in Fig 2 and 3 suggest the contours become ellipses. So scaling a_{11} (or other diagonal entries) has the effect of compressing that direction. If $A = \begin{bmatrix} 1, 0 \\ 0, 4 \end{bmatrix}$, the effect is compressing x_2 direction.

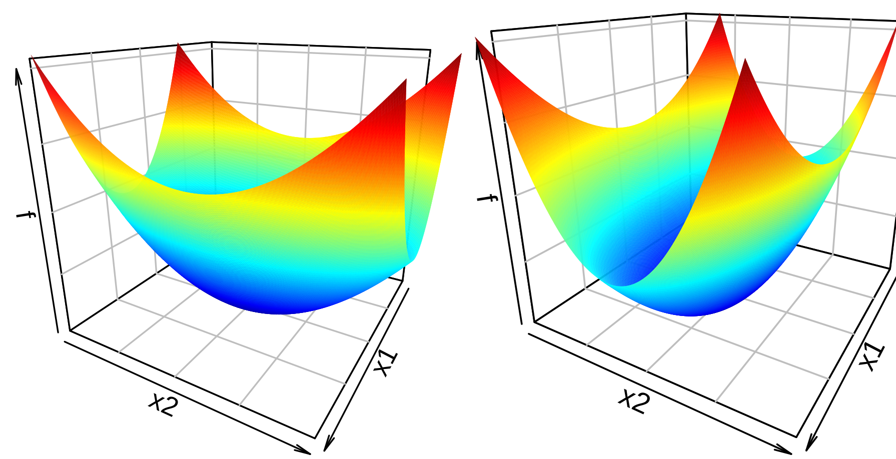

Figure 2: surface plot of two axis-aligned elliptic paraboloid: scaling diagonal entry of A has the effect of compressing the bowl

Example 3.

$$A = \begin{bmatrix} 3, 4 \\ 4, 3 \end{bmatrix}$$

The plots shown in Fig 4 suggest the contours become rotated ellipses. Therefore,

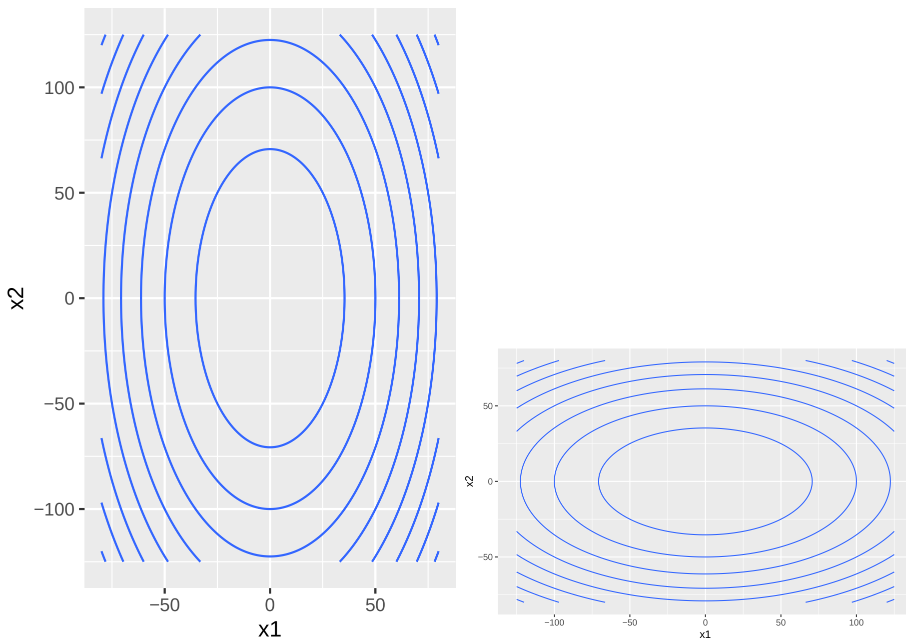

Figure 3: contour plots of two axis aligned elliptic paraboloid

non-diagnonal entries of A rotate the elliptic paraboloid (A has to be either positive definite or negative definite).

2 Common quadratic forms in machine learning models

Quadatric form, $x^T Ax$, is used a lot in machine learning models. The following are three cases.

Linear regression: the loss function of a linear regression is

$$L(\theta) = (y - X\theta)^T (y - X\theta),$$

which is a quadratic form of θ.

Gaussian distribution: the kernel of a multivariate Gaussian is a quadratic form, namely

$$-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu),$$

where μ, Σ are the mean and variance-covariance matrix respectively. The kernel is a quadratic form of input x.

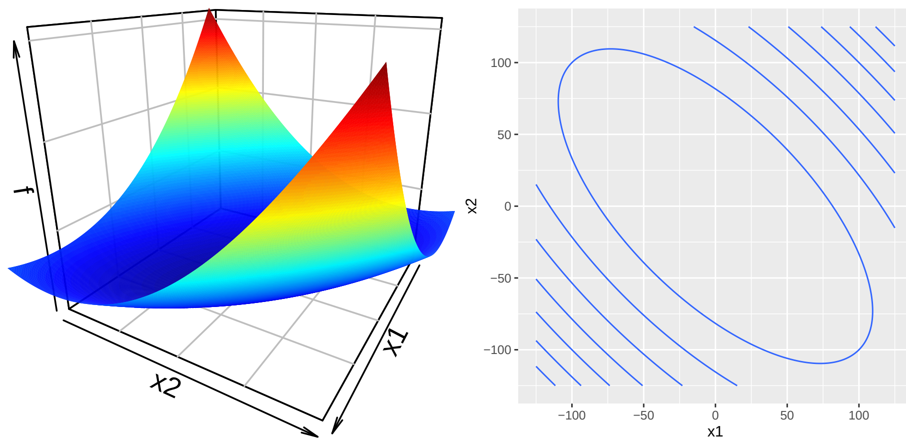

Figure 4: surface plot of a rotated elliptic paraboloid: non-diagonal entries rotate the ellipse

Taylor expansion: Taylor's expansion contains quadratic forms as well:

$$T(x) = f(a) + \nabla_x f(a)(x-a) + \frac{1}{2!}\underbrace{(x-a)^T \nabla_x^2 f(a)(x-a)}_{q.f.} + \dots,$$

where $\nabla_x f(a)$ is the gradient of f and $\nabla_x^2 f(a)$ is the hessian matrix. Note that we define gradients as row vectors, so the second term $\nabla_x f(a)(x-a)$ is an inner product. The expansion approximates a multivariate $R^n \rightarrow R$ function $f(x)$ by a polynomial function $T(x)$. The third term is a quadratic form of input x.

3 Gradient and Hessian of a quadratic form

It should be easy to remember the following results if you compare it with univariate quadratic function $f(x) = ax^2 = xax$, whose gradient is $f' = 2ax$ and second order derivative is $f'' = 2a$.

3.1 Gradient of a quadratic form

The gradient of $f(x) = x^T Ax$ is

$$\nabla_x f(x) = x^T(A + A^T)$$

Note that as we have adopted the convention that gradients are row vectors, hence the gradient is written as $2x^T A$, a $1 \times n$ vector.

If A is symmetric, then $A = A^T$, therefore

$$\nabla_x f(x) = x^T(A + A^T) = 2x^T A$$

3.2 Hessian of a quadratic form

The Hessian or gradient of the gradient of $f(x) = x^T Ax$ is

$$\nabla_x^2 f(x) = A^T + A$$

If A is symmetric, the Hessian is

$$\nabla_x^2 f(x) = 2A$$

3.3 Maximum and minimum of a quadratic form

- If $\nabla_x^2 f(x)$ is positive definite, $f(x) = x^T Ax$ has a minimum;
- If $\nabla_x^2 f(x)$ is negative definite, $f(x) = x^T Ax$ has a maximum;

3.3.1 Positive definite matrix

If matrix A is positive definite, then $x^T Ax > 0$ for $\forall x \in R^n$

4 Positive definite quadratic form

a multivaraite Gaussian distribution (k dimensional) is

$$p(x) = (2\pi)^{-\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

This note lists some important facts about quadratic form.

First and foremost, it returns a scalar given an input of vector x.

$$x^T Ax : R^m \to R$$