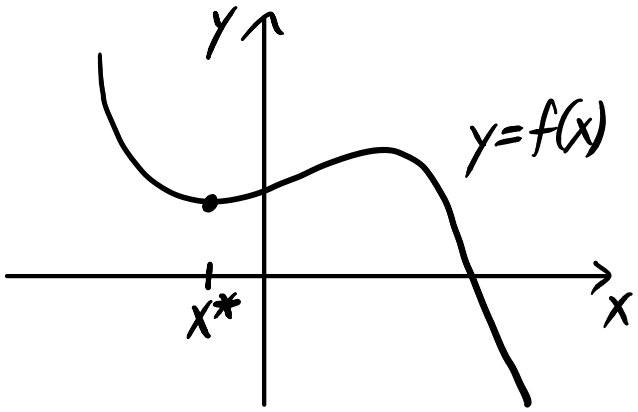
I.7 | Symmetric Positive Definite Matrices

Motivation: Optimization



$\boxed{x^* \text{ is a local minimizer of } f}$

From calculus we know that a twice diff'ble function $f: \mathbb{R} \to \mathbb{R}$ has a local minimizer at x^* if:

(1)

(2)

Training a neural network involves minimizing a "loss function" that has hundreds or thousands of variables. The loss function measures the error.

Consider minimizing a twice diff'ble function $f: \mathbb{R}^n \to \mathbb{R}$ having n variables. In this case, x^* is a <u>local minimizer</u> of f if:

(1) $\nabla f(x^*) = 0$

(2) $\nabla^2 f(x^*)$ is positive definite

Here $\nabla f(x)$ is the <u>gradient</u> of f at x,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix} \in \mathbb{R}^n,$$

and $\nabla^2 f(x)$ is the <u>Hessian</u> of f at x,

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix} \in \mathbb{R}^{n \times n}.$$

Note that since $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$,

the Hessian $\nabla^2 f(x)$ is a symmetric matrix.

Just like how $f''(x)$ tells us if the graph of f is <u>concave up</u> or <u>concave down</u> at x, the eigenvalues of $\nabla^2 f(x)$ tell us if f is <u>convex</u> (<u>ie</u> concave up), <u>concave</u> (<u>ie</u> concave down), or <u>neither</u> at x.

<u>Example</u>:

$f(x,y) = x^4 - x^2 + y^2$ [credit: Khan Academy]