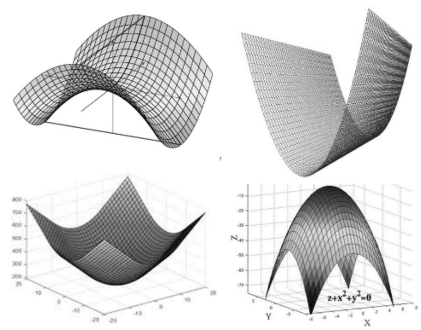
Example: Which one is convex?



Notes

Verifying Convexity

Any of the following conditions is **necessary** and **sufficient** for convexity:

1. By definition:

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y}).$$

   - Function value is lower than the line.

2. First Order Convexity:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}), \qquad \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}.$$

   - Tangent line is always lower than the function

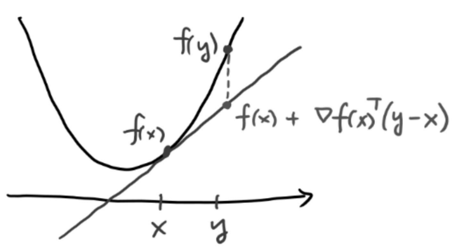
3. Second Order Convexity: f is convex over $\mathcal{X}$ if and only if

$$\nabla^2 f(\mathbf{x}) \succeq 0 \quad \forall \mathbf{x} \in \mathcal{X}.$$

   - Curvature is positive.

Notes

Tangent Line Condition Illustrated



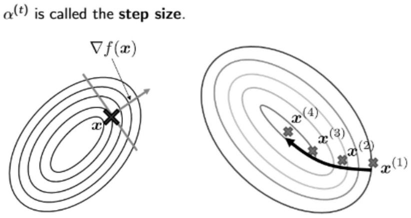
Notes

Gradient Descent

The algorithm:

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \alpha^{(t)} \nabla f(\mathbf{x}^{(t)}), \qquad t = 0, 1, 2, \ldots,$$

where $\alpha^{(t)}$ is called the **step size**.

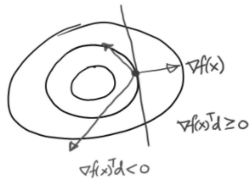


Notes

Gradient Descent

Pictorial illustration:

- $\nabla f(\boldsymbol{x})$ is **perpendicular** to the contour.
- A search direction $\boldsymbol{d}$ can either be on the positive side $\nabla f(\boldsymbol{x})^T\boldsymbol{d} \geq 0$ or negative side $\nabla f(\boldsymbol{x})^T\boldsymbol{d} < 0$.
- Only those on the negative side can reduce the cost.
- All such $\boldsymbol{d}$'s are called the **descent directions**.



Notes

Gradient Descent

$$f'(X)=\begin{bmatrix}8x_1+3x_2-5.5\\3x_1+5x_2-4\end{bmatrix}$$

Learning parameter (α) = 0.135

Initial guess $(X_0)=\begin{bmatrix}2\\2\end{bmatrix}$ $\quad f(X_0)=19$

Step 1: $X_1=X_0-\alpha f'(X_0)$ ← Gradient Descent (or) Learning Rule in ML

$$X_1=\begin{bmatrix}2\\2\end{bmatrix}-0.135\begin{bmatrix}8x_{0,1}+3x_{0,2}-5.5\\3x_{0,1}+5x_{0,2}-4\end{bmatrix}$$

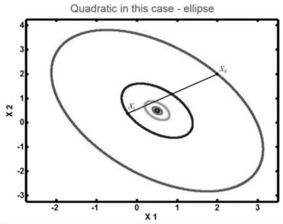
$$X_1=\begin{bmatrix}2\\2\end{bmatrix}-0.135\begin{bmatrix}8(2)+3(2)-5.5\\3(2)+5(2)-4\end{bmatrix}$$

$$X_1=\begin{bmatrix}-0.2275\\0.3800\end{bmatrix}\quad f(X_1)=0.0399$$

Constant objective function contour plots

$$f(X)=4x_1^2+3x_1x_2+2.5x_2^2-5.5x_1-4x_2=K$$

Quadratic in this case - ellipse



Notes

Gradient Descent

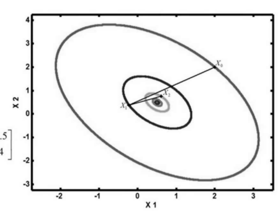
First iteration $(X_1)=\begin{bmatrix}-0.2275\\0.3800\end{bmatrix}$

Step 2: $X_2=X_1-\alpha f'(X_1)$

$$X_2=\begin{bmatrix}-0.2275\\0.3800\end{bmatrix}-0.135\begin{bmatrix}8x_{1,1}+3x_{1,2}-5.5\\3x_{1,1}+5x_{1,2}-4\end{bmatrix}$$

$$X_2=\begin{bmatrix}-0.2275\\0.3800\end{bmatrix}-0.135\begin{bmatrix}8(-0.2275)+3(0.3800)-5.5\\3(-0.2275)+5(0.3800)-4\end{bmatrix}$$

$$X_2=\begin{bmatrix}0.6068\\0.7556\end{bmatrix}\quad f(X_2)=-2.0841$$



Notes

Gradient Descent

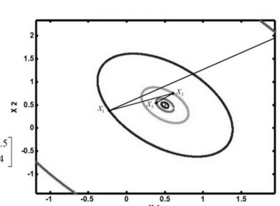
Second iteration $(X_2)=\begin{bmatrix}0.6068\\0.7556\end{bmatrix}$

Step 3: $X_3=X_2-\alpha f'(X_2)$

$$X_3=\begin{bmatrix}0.6068\\0.7556\end{bmatrix}-0.135\begin{bmatrix}8x_{2,1}+3x_{2,2}-5.5\\3x_{2,1}+5x_{2,2}-4\end{bmatrix}$$

$$X_3=\begin{bmatrix}0.6068\\0.7556\end{bmatrix}-0.135\begin{bmatrix}8(0.6068)+3(0.7556)-5.5\\3(0.6068)+5(0.7556)-4\end{bmatrix}$$

$$X_3=\begin{bmatrix}0.3879\\0.5398\end{bmatrix}\quad f(X_3)=-2.3342$$



Notes

Gradient Descent

Third iteration $(X_3) = \begin{bmatrix} 0.3879 \\ 0.5398 \end{bmatrix}$

Step 4: $X_4 = X_3 - \alpha f'(X_3)$

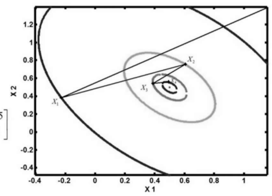
$$X_4 = \begin{bmatrix} 0.3879 \\ 0.5398 \end{bmatrix} - 0.135 \begin{bmatrix} 8x_{3,1} + 3x_{3,2} - 5.5 \\ 3x_{3,1} + 5x_{3,2} - 4 \end{bmatrix}$$

$$X_4 = \begin{bmatrix} 0.3879 \\ 0.5398 \end{bmatrix} - 0.135 \begin{bmatrix} 8(0.3879) + 3(0.5398) - 5.5 \\ 3(0.3879) + 5(0.5398) - 4 \end{bmatrix}$$

$$X_4 = \begin{bmatrix} 0.4928 \\ 0.5583 \end{bmatrix} \qquad f(X_4) = -2.3675$$

Optimal solution $(X_{opti}) = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$ $\quad f(X_{opti}) = -2.3750$

Gradient is zero at the optimum point



Notes

Notes

Notes

Notes