1. Directional Derivatives.

2. Motivation for the topic Lagrange multiplier

(a) Application of Lagrange multipliers in classification algorithms

↳ Support vector machine (SVM)

EE1203: Vector Calculus

Aneesh Sobhanan
Department of Electrical Engineering
IIT Hyderabad, India

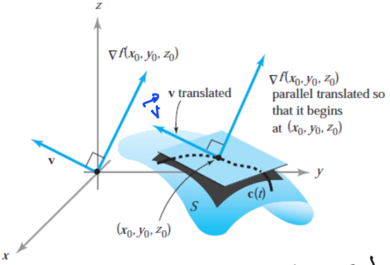
23rd January 2025.

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

1. <u>Theorem</u> : $\nabla f_{(x_0,y_0,z_0)}$ is $\perp^r$ to the level surface ($\perp^r$ to tangent plane at (x_0, y_0, z_0))

or ∇f is orthogonal to the <u>surface 'S'</u> <u>on which f is constant</u> $\rightarrow$ level surface.

figure 2.6.2 Geometric significance of the gradient: ∇f is orthogonal to the surface S on which f is constant.



$S \rightarrow f = \text{constant}.$

$c(t) = \vec{r} + t\vec{v}$

From : Ref: Vector Calculus 6^{th} Ed

By J.E Marsden
A. Tromba.

The rate of change of function 'f' as you move along curve $c(t)$ (By chain rule)

$$\left.\frac{df}{dt}\right|_{(x_0,y_0,z_0)} = \left.\nabla f\right|_{(x_0,y_0,z_0)} \cdot \frac{d\bar{r}}{dt}$$

velocity vector

$\bar{v} = \frac{d\bar{r}}{dt} \rightarrow$ tangent of point (x_0, y_0, z_0)

$$= \left.\nabla f \cdot \vec{v}\right|_{(x_0,y_0,z_0)}$$

* tangent vector $\vec{v}$ is tangent to the surface S too as the curve $c(t)$ is inside the surface 'S'.

ie, $\vec{v}$ is tangent to the level surface $f = \text{Const.}$

$\frac{df}{dt} = 0 \rightarrow$ in 'S' (level surface)

$\Rightarrow \nabla f \cdot \vec{v} = 0$

$\Rightarrow$ Gradient vector is $\perp^r$ to the tangent vector

$\Rightarrow \perp^r$ to the level surface.

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Directional derivatives.

$W = W(x,y) \rightsquigarrow$ $\frac{\partial w}{\partial x} \to$ Change in 'w' in 'x' direction.

$\frac{\partial w}{\partial y} \to$ Change in 'y' direction.

$\Rightarrow \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}$ are derivatives in direction of $\hat{i}$ or $\hat{j}$.

Rate of change of function $w(x,y)$ in any random direction $\vec{v}$ is;

From the understanding of the chain rule again;

$$\frac{dw}{dt} = \nabla w \cdot \frac{d\vec{r}}{dt}$$

$$= \nabla w \cdot \vec{v}.$$

Rate of change of 'w' along direction $\vec{v}$ is

$$\frac{dw}{dt} = \nabla w \cdot \vec{v}.$$

But directional derivative of 'w' along '$\vec{v}$' is

$$\frac{dw}{dt} = \nabla w \cdot \hat{v}$$

; $\hat{v} = \frac{\vec{v}}{|\vec{v}|}$ → unit vector.

why unit vector?

* More generally, the definition of directional derivative is true measure of rate of change of $w(x,y,z)$ w.r.t distance along a curve (path) in given direction if $\vec{u}$ is normalized to $\hat{u}$.

For easy understanding consider the following:

directional derivative along $\hat{i}$

$$= \nabla w \cdot \hat{i}$$

$$= \frac{\partial w}{\partial x} \checkmark$$

→ LAGRANGE multipliers: Optimization problem with bounds/ Constraints.

Hyperplane:

In R^2 : Hyperplane is a line.

In R^3 : Hyperplane is a plane

In R^n : Hyperplane is a $(n-1)$ dimensional subspace.

Hyperplane in R^n is given by;

$$a_1 x_1 + a_2 x_2 + \dots a_n x_n = b.$$

$x_1, x_2, \dots x_n$: co-ordinates of a point on the hyperplane.

$a_1, a_2, \dots a_n$: Constants that define the orientation of the hyperplane.

b : The constant that determines the position of hyperplane relative to origin.

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

* Optimal separating Hyperplane:

Let us operate in R^3

Hyperplane in R^3 ; $H = \{ \bar{x} : \vec{m}\cdot\bar{x} + c = 0 \}$ $\bar{x} \in R^3$

$\bar{x} = [x_1, x_2 \; x_3]^T$; $\bar{m} = [m_1, m_2, m_3]$; $c \in \mathbb{R}$

Let $\bar{x}_a, \bar{x}_b \in H$ then $\vec{m}\cdot(\bar{x}_a - \bar{x}_b) = 0$

$\Rightarrow \vec{m} \perp H$

ie; $\bar{m}$ is the normal vector to the Hyperplane H.

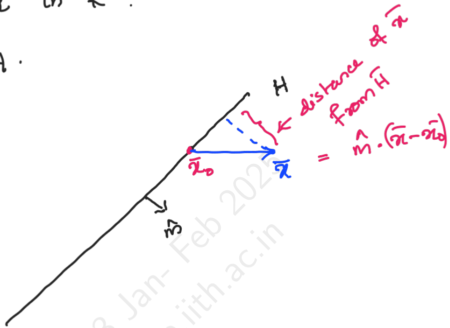
$\hat{m} = \dfrac{\vec{m}}{|\vec{m}|}$ unit normal to H.

* If $\bar{x}_0 \in H \implies \bar{m}\cdot\bar{x}_0 + C = 0$

Let $\bar{x}$ be any point in $\mathbb{R}^3$.

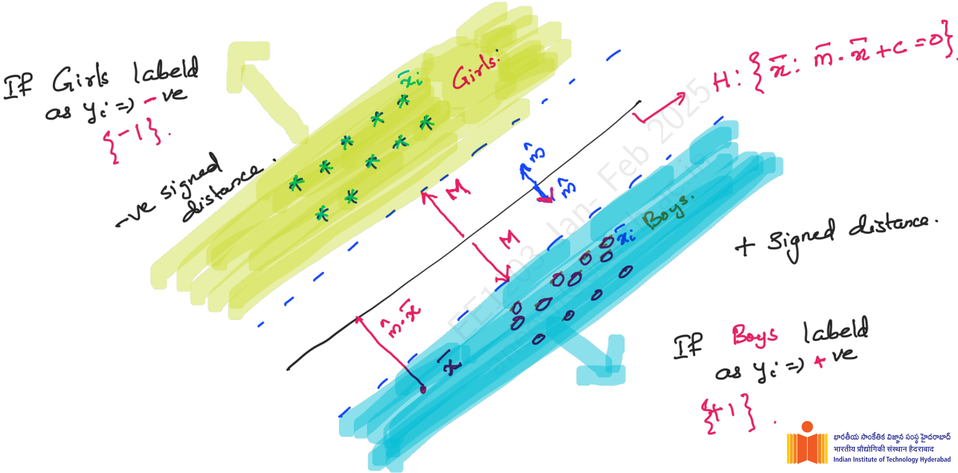
How far is $\bar{x}$ from H.

$\implies \hat{m}\cdot(\bar{x}-\bar{x}_0)$



Application: Optimal separating Hyperplane in algorithm like "Support vector machine". (SVM) : For classification problems.

Problem: Given two classes (say Boys and Girls) via their attributes, find the optimal separating Hyperplane:



* Let us take $\mathbb{R}^3$ space

* $H = \{ \bar{x} : \bar{m} \cdot \bar{x} + c = 0 \}$ Hyperplane.

Optimal Hyperplane : It means the Hyperplane should be farthest possible away from the two classes (Here, boys & Girls) at the same time distinctly separating them.

* If $\bar{x}_a, \vec{x}_b \in H$ $\quad \bar{m} \cdot (\bar{x}_a - \bar{x}_b) = 0$; $\bar{m}$ is normal to H ---(1)

* $\hat{m} = \dfrac{\bar{m}}{|\bar{m}|}$ unit normal to H ---(2)

* For any $\bar{x}_0 \in H$, $\bar{m}\cdot\bar{x}_0 + c = 0$ or $\bar{m}\cdot\bar{x}_0 = -c$ ----- (3)

* The signed distance of any $\bar{x} \in \mathbb{R}^3$ to H is given by projection of $(\bar{x}-\bar{x}_0)$ onto $\hat{m}$.

$$\hat{m}\cdot(\bar{x}-\bar{x}_0) = \hat{m}\cdot\bar{x} - \hat{m}\cdot\bar{x}_0$$

$$= \hat{m}\cdot\bar{x} - \frac{\bar{m}\cdot\bar{x}_0}{|\bar{m}|}$$

$$= \hat{m}\cdot\bar{x} + \frac{c}{|\bar{m}|} \quad \because \text{From Eqn(3)}$$

$$= \underbrace{\frac{\bar{m}\cdot\bar{x} + c}{|\bar{m}|}}_{\text{signed distance of } \bar{x} \text{ from } H} \quad \text{---} (*)$$

Problem: Find $(\bar{m}, c)$ such that the product of signed distance of a point (Girl/boy) with its label is as high as possible ie;

if $\bar{x}_1, \bar{x}_2, \cdots \bar{x}_{100}$ are the points corresponding to boys & Girls (100 in numbers) then H should be such that

$$f(\bar{x}_i, y_i, \bar{m}) = \sum_{i=1}^{100} \underset{\substack{\text{label} \\ (+1/-1)}}{y_i} \; \underbrace{\frac{\bar{m}\cdot\bar{x}_i + c}{|\bar{m}|}}_{\text{signed distance}} \quad \text{is maximized.} \quad \text{---}(5)$$

y_i: label given to data points. $\bar{x}_i$

* We can reformulate Eqn (5) using a barrier value M such that

$$y_i\left(\frac{\bar{m}\cdot\bar{x}_i + c}{|\bar{m}|}\right) \geq M \quad \text{for all } i, \quad \text{---}(6)$$

Problem: Find the parameters $(\vec{m}, c)$ such that M is as large as possible.

define optimal H.

* For a new data point $\bar{x}_{new}$ simply find

$$\left(\frac{\vec{m}\cdot\bar{x}_{new} + c}{|\vec{m}|}\right) (\Rightarrow \text{signed distance})$$

If this signed distance is > 0 then $\bar{x}_{new}$ is Boy

If this signed distance is ≤ 0 then $\bar{x}_{new}$ is Girl.