A Short Outline of Lagrange Multipliers.

Example. Find a point $\vec{p}$ on the plane

$$x+y-2z=6$$

which lies closest to the origin.

→: Q: Why is this a constrained optimization problem?

→: Function that we want to minimize

$$\tilde{D}(x,y,z)=x^2+y^2+z^2$$

subject to : $\boxed{x+y-2z=6}$

11/18/2024.

In general, $f(x,y,z) \longrightarrow \min/\max$

subject to the constraint $F(x,y,z)=0$

First, we construct the "Lagrangian function"

$$L(x,y,z,\lambda)=f(x,y,z)+\lambda F(x,y,z)$$

Then, we optimize the function L as a function of four variables (x,y,z,λ)

Back to our example:

$\tilde{D}(x,y,z)=x^2+y^2+z^2 \longrightarrow \min$

subject to $F(x,y,z)=x+y-2z-6=0$

$\Rightarrow L(x,y,z,\lambda)=x^2+y^2+z^2+\lambda(x+y-2z-6)$

$$\left.\begin{aligned}
\frac{\partial L}{\partial x} &= 2x+\lambda = 0 &&\Rightarrow x=-\frac{\lambda}{2}\\
\frac{\partial L}{\partial y} &= 2y+\lambda = 0 &&\Rightarrow y=-\frac{\lambda}{2}\\
\frac{\partial L}{\partial z} &= 2z-2\lambda = 0 &&\Rightarrow z=\lambda
\end{aligned}\right\} \vec{p}=(x,y,z)$$

$$\frac{\partial L}{\partial \lambda} = x+y-2z-6 = 0 \quad\Rightarrow\quad x+y-2z-6=0$$

$$\Rightarrow -\frac{\lambda}{2} - \frac{\lambda}{2} - 2\lambda = 6$$

$$\Rightarrow \boxed{\lambda = -2}$$

$$\Rightarrow \boxed{\vec{\beta} = (1, 1, -2)}$$

$\square$

Margins & Separating Hyperplanes.

Linear classifiers can be described geometrically as separating hyperplanes

Any affine function $x \longmapsto \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$ determines a hyperplane in $\mathbb{R}^p$ our perdictor space

More precisely, $\{x : \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p = 0\}$ is a hyperplane splitting the space $\mathbb{R}^p$ into two "half spaces"

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p > 0$$

and

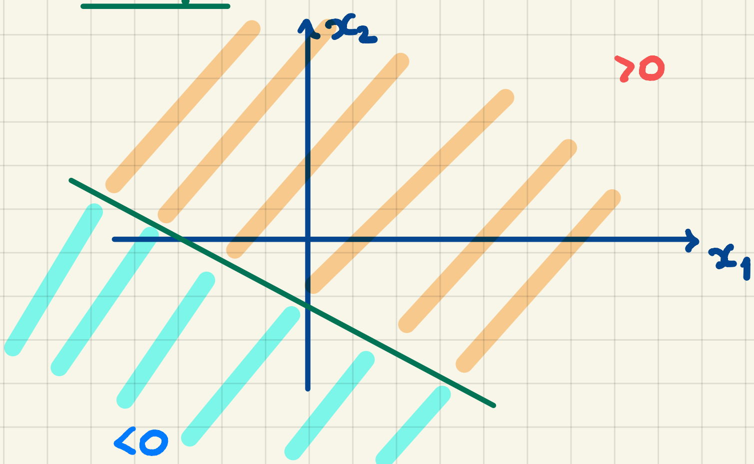
$$\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p < 0$$

The vector $\vec{n} = (\beta_1, \beta_2, \ldots, \beta_p)$ is the normal vector of our hyperplane. For a given hyperplane, we can always choose $\vec{n}$ so that

$$\|\vec{n}\| = 1$$

Of course, the coefficient β_0 must also be scaled.

Example. Consider $x \longmapsto 1 + 2x_1 + 3x_2$



Note. • If the hyperplane goes through the origin, then $\beta_0 = 0$

For any point $x = (x_1, \dots, x_p)$ in the space, the deviation between it and the hyperplane w/ $\beta = (\beta_1, \dots, \beta_p)$ is equal to

$$x \cdot \beta = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

• If $\beta_0 \neq 0$, the hyperplane does not go through the origin. The deviation becomes

$$\boxed{\beta_0 + x \cdot \beta}$$

The sign tells us which side of the hyperplane we're.

Maximal Margin Classifier.

Suppose that we have a classification problem w/ two classes. We choose to encode these classes as $Y = -1$ and $Y = +1$.

Our criterion for the best among all the separating hyperplanes (if such exist) is to find the one the largest possible margin around the hyperplane.

$\boxed{\text{OPTIMIZATION PROBLEM.}}$