

Read [Turk and Pentland](#).

## Linear Discriminant Functions

Line is like a 2-class classifier

Put point in line equation

- positive value: 1st class
- negative values: 2nd class

## Gradient Descent

We iteratively modify  $w_i$  so that each sample is correctly classified.

We are gonna define loss function  $J(w)$ , which will be minimum when the training samples are correctly classified.

We are creating another space  $J(w)$  vs  $w_1$  vs  $w_2$  etc

We want to minimize  $J(w)$

**What modifications will reduce  $J(w)$ ?**

- Compute partial derivatives
- $\frac{\partial J}{\partial w}$  will be a vector of all partial derivatives.
- Gradient Vector:

$$\nabla J = \frac{\partial J}{\partial w}$$

- $\eta$ : Learning rate

## Gradient Descent Algorithm