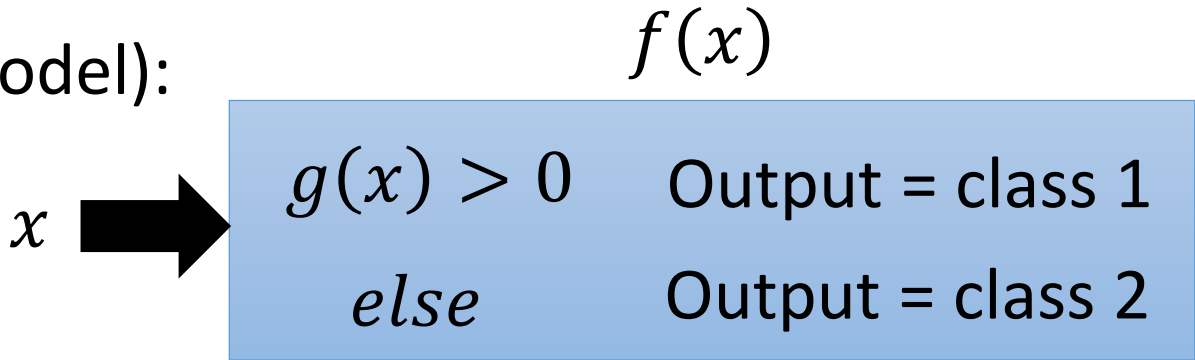
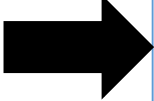


# Classification: Probabilistic Generative Model

# Ideal Alternatives

- Function (Model):  


$x$    $f(x)$

$g(x) > 0$	Output = class 1
<i>else</i>	Output = class 2

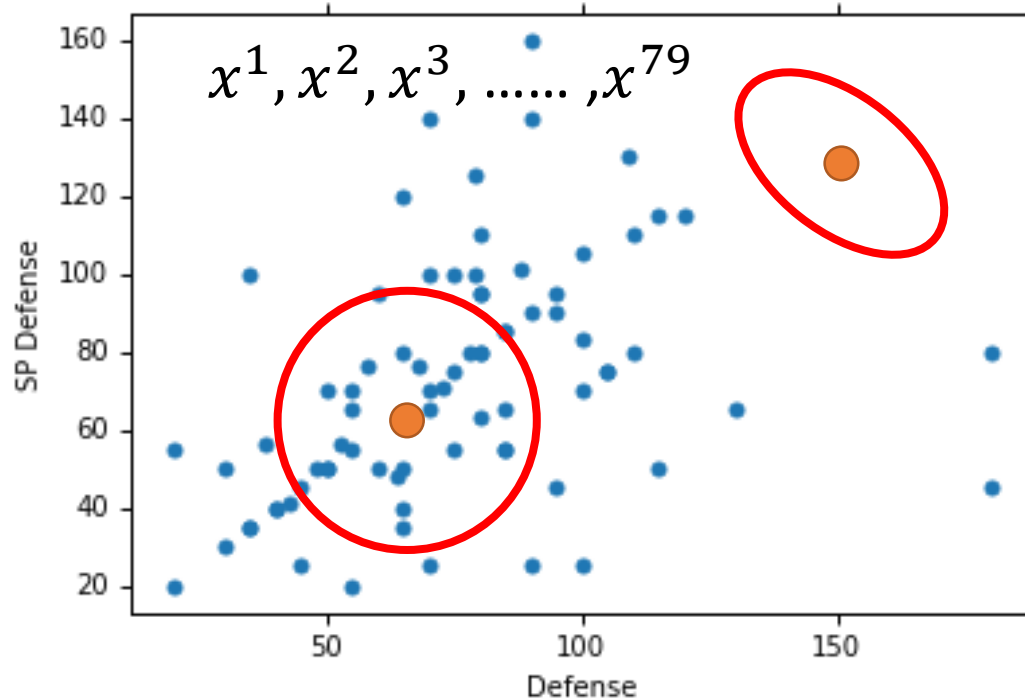
- Loss function:

$$L(f) = \sum_n \delta(f(x^n) \neq \hat{y}^n)$$

The number of times  $f$  get incorrect results on training data.

- Find the best function:
  - Example: Perceptron, SVM

**Maximum Likelihood**  $f_{\mu, \Sigma}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$



The Gaussian with any mean  $\mu$  and covariance matrix  $\Sigma$  can generate these points. ➡ Different Likelihood

Likelihood of a Gaussian with mean  $\mu$  and covariance matrix  $\Sigma$   
 = the probability of the Gaussian samples  $x^1, x^2, x^3, \dots, x^{79}$

$$L(\mu, \Sigma) = f_{\mu, \Sigma}(x^1) f_{\mu, \Sigma}(x^2) f_{\mu, \Sigma}(x^3) \dots f_{\mu, \Sigma}(x^{79})$$

# Maximum Likelihood

We have the “Water” type Pokémons:  $x^1, x^2, x^3, \dots, x^{79}$

We assume  $x^1, x^2, x^3, \dots, x^{79}$  generate from the Gaussian  $(\mu^*, \Sigma^*)$  with the **maximum likelihood**

$$L(\mu, \Sigma) = f_{\mu, \Sigma}(x^1) f_{\mu, \Sigma}(x^2) f_{\mu, \Sigma}(x^3) \dots f_{\mu, \Sigma}(x^{79})$$

$$f_{\mu, \Sigma}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}$$

$$\mu^*, \Sigma^* = \arg \max_{\mu, \Sigma} L(\mu, \Sigma)$$

$$\mu^* = \frac{1}{79} \sum_{n=1}^{79} x^n$$

average

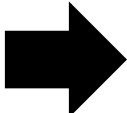
$$\Sigma^* = \frac{1}{79} \sum_{n=1}^{79} (x^n - \mu^*) (x^n - \mu^*)^T$$

# Now we can do classification

$$f_{\mu^1, \Sigma^1}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma^1|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu^1)^T (\Sigma^1)^{-1} (x - \mu^1)\right\}$$

$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$

$$f_{\mu^2, \Sigma^2}(x) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma^2|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu^2)^T (\Sigma^2)^{-1} (x - \mu^2)\right\}$$

If  $P(C_1|x) > 0.5$   x belongs to class 1 (Water)

Modifying Model  The same  $\Sigma$   
Less parameters

# Three Steps

- Function Set (Model):

$x$  

$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$

If  $P(C_1|x) > 0.5$ , output: class 1  
Otherwise, output: class 2

- Goodness of a function:
  - The mean  $\mu$  and covariance  $\Sigma$  that maximizing the likelihood (the probability of generating data)
- Find the best function: easy

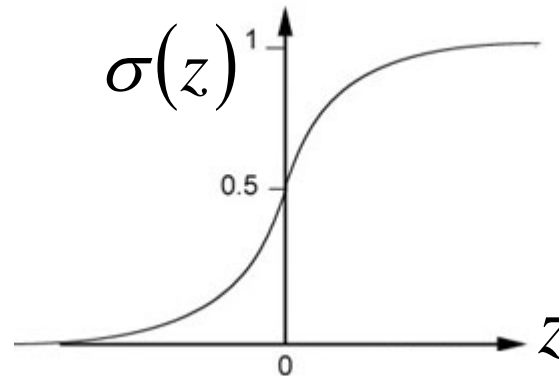
# Sigmoid function

$$P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)}$$

$$= \frac{1}{1 + \frac{P(x|C_2)P(C_2)}{P(x|C_1)P(C_1)}} = \frac{1}{1 + \exp(-z)} = \sigma(z)$$

Sigmoid function

$$z = \ln \frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)}$$



$$P(C_1|x) = \sigma(z)$$

$$\Sigma_1 = \Sigma_2 = \Sigma$$

$$z = \underbrace{(\mu^1 - \mu^2)^T \Sigma^{-1} x}_{\mathbf{w}^T} - \underbrace{\frac{1}{2} (\mu^1)^T \Sigma^{-1} \mu^1 + \frac{1}{2} (\mu^2)^T \Sigma^{-1} \mu^2 + \ln \frac{N_1}{N_2}}_{b}$$

$$P(C_1|x) = \sigma(\mathbf{w} \cdot x + b)$$

How about directly find  $\mathbf{w}$  and  $b$ ?

In generative model, we estimate  $N_1, N_2, \mu^1, \mu^2, \Sigma$

Then we have  $\mathbf{w}$  and  $b$