1. Normal or Gaussian p.d.f., $N(\mu, \sigma^2)$,

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e\left[\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

2. ~~Entropy~~ Entropy of a p.d.f $p(x)$ is given by

$$H(p(x)) = -\int p(x) \ln p(x) \, dx$$

3. Central Limit Theorem: The aggregate effect of ~~the~~ the sum of a large number of small i.i.d(s) tend to a normal distribution.

4. For a given (μ, σ), normal distribution has

the maximum entropy.

5. Normal multivariate:

$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} e\left[-\frac{1}{2} (x-\mu)' \Sigma^{-1} (x-\mu)\right]$$

where Σ is the covariance matrix

$$\Sigma = E\left[(x-\mu)(x-\mu)'\right] = \int (x-\mu)(x-\mu)' p(x) \, dx$$

The diagonal elements of Σ are σ_{ii} i.e σ_i^2, variances and off-diagonal element σ_{ij} are the covariances of x_i and x_j.

Discriminant fns for Normal p.d.f.

$$g_i(x) = \ln p(x|\omega_i) + \ln P(\omega_i)$$

$p(x|\omega_i)$ is $N(\mu_i, \Sigma_i)$ then

$$g_i(x) = -\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i).$$

Case 1

$\Sigma_i = \sigma^2 I$, the features are statistically independent and each feature has the same variance, σ^2.

Then $g_i(x) = -\frac{\|x-\mu_i\|^2}{2\sigma^2} + \ln P(\omega_i)$

But, $\|x-\mu_i\|^2 = (x-\mu_i)'(x-\mu_i)$, so

$$g_i(x) = -\frac{1}{2\sigma^2}\left[x'x - 2\mu_i'x + \mu_i'\mu_i\right] + \ln P(\omega_i).$$

Since $x'x$ is same for all i, ignoring it, an equivalent linear discriminant function is ~~$g_i(x)$~~ $g_i(x) = T_i'x + T_{i0}$

where $T_i = \frac{1}{\sigma^2}\mu_i$ and

$$T_{i0} = \frac{-1}{2\sigma^2}\mu_i'\mu + \ln P(\omega_i)$$

This classifier is called a linear machine.