Parametric Probability Models

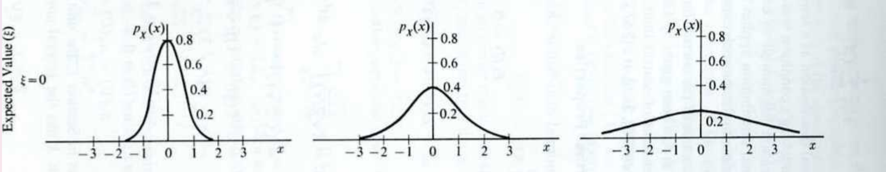
Robert M. Haralick

Computer Science, Graduate Center
City University of New York

Outline

The Multivariate Normal Density Function

$$p(x) = \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu))$$



The Density Function

$$I = \int_x \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu))dx$$

Let $y = x - \mu$

$$I = \int_y \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}y'\Sigma^{-1}y)dy$$

The Density Function

Now by the eigenvalue eigenvector decomposition of Σ, $\Sigma = T'DT$, where T is orthonormal and D is diagonal. Hence,

$$\begin{aligned}\Sigma^{-1} &= (T'DT)^{-1} \\ &= T^{-1}D^{-1}(T')^{-1} \\ &= T'D^{-1}T \\ &= T'D'^{-\frac{1}{2}}D^{-\frac{1}{2}}T\end{aligned}$$

so that

$$\begin{aligned}I &= \int_y \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\exp(-y'T'D'^{-\frac{1}{2}}D^{-\frac{1}{2}}Tydy \\ &= \int_y \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\exp(-(D^{-\frac{1}{2}}Ty)'(D^{-\frac{1}{2}}Ty))dy\end{aligned}$$

The Density Function

$$I = \int_y \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-(D^{-\frac{1}{2}}Ty)'(D^{-\frac{1}{2}}Ty))dy$$

Let $z = D^{-\frac{1}{2}}Ty$ so that $y = T'D^{\frac{1}{2}}z$. The Jacobian is $\frac{\partial y}{\partial z}$.

$$\begin{aligned} \frac{\partial y}{\partial z} &= T'D^{\frac{1}{2}} \\ |\frac{\partial y}{\partial z}| &= |T'D^{\frac{1}{2}}| \\ &= |T'|\,|D^{\frac{1}{2}}| \\ &= |D|^{\frac{1}{2}} \end{aligned}$$

The Density Function

$$I = \int_y \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-(D^{-\frac{1}{2}}Ty)'(D^{-\frac{1}{2}}Ty))dy$$

Hence,

$$I = \int_z \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-z'z)|D|^{\frac{1}{2}}dz$$

The Density Function

Recall,

$$
\begin{aligned}
\Sigma &= T'DT \\
|\Sigma| &= |T'DT| \\
&= |T'|\,|D|\,|T| \\
&= |D|
\end{aligned}
$$

Hence, $|\Sigma|^{\frac{1}{2}} = |D|^{\frac{1}{2}}$.

The Density Function

$$
\begin{aligned}
I &= \int_z \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-z'z)|D|^{\frac{1}{2}} dz \\
&= \int_z \frac{1}{|2\pi|^{\frac{N}{2}}} \exp(-z'z) dz \\
&= \prod_{n=1}^{N} \int_{z_n} \frac{1}{(2\pi)^{\frac{1}{2}}} \exp(-z_n^2) dz_n \\
&= \prod_{n=1}^{N} 1 \\
&= 1
\end{aligned}
$$

The Mean

$$
\begin{aligned}
p(x) &= \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)) \\
E[x] &= \int_x xp(x)dx \\
&= \int_x \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} x \exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu))dx
\end{aligned}
$$

The Mean

Let $y = x - \mu$.
Note that odd functions integrated over even limits result in zero.

$$\begin{aligned}
E[x] &= \int_y \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}(y+\mu)\exp(-\frac{1}{2}y'\Sigma^{-1}y)dy \\
&= \int_y \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}y\exp(-\frac{1}{2}y'\Sigma^{-1}y)dy + \\
&\qquad \int_y \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\mu\exp(-\frac{1}{2}y'\Sigma^{-1}y)dy \\
&= 0 + \mu\int_y \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\exp(-\frac{1}{2}y'\Sigma^{-1}y)dy \\
&= \mu
\end{aligned}$$

The Covariance

$$
\begin{aligned}
Cov(x) &= E[(x-\mu)(x-\mu)'] \\
&= \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\int_x (x-\mu)(x-\mu)' \exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu))dx
\end{aligned}
$$

Let $y = x - \mu$

$$
Cov(x) = \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\int_y yy' \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}y'\Sigma^{-1}y)dy
$$

The Covariance

Let $z = D^{-\frac{1}{2}}Ty$. Then $y = T'D^{\frac{1}{2}}z$ so that

$$
\begin{aligned}
Cov(x) &= \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\int_z T'D^{\frac{1}{2}}z(T'D^{\frac{1}{2}}z)'\exp(-\frac{1}{2}z'z)|\Sigma|^{\frac{1}{2}}dz \\
&= \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\int_z T'D^{\frac{1}{2}}zz'D^{\frac{1}{2}}T\exp(-\frac{1}{2}z'z)|\Sigma|^{\frac{1}{2}}dz \\
&= T'D^{\frac{1}{2}}\left(\frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}}\int_z zz'\exp(-\frac{1}{2}z'z)|\Sigma|^{\frac{1}{2}}dz\right)D^{\frac{1}{2}}T \\
&= \frac{1}{|2\pi|^{\frac{N}{2}}}T'D^{\frac{1}{2}}\left(\int_z zz'\exp(-\frac{1}{2}z'z)dz\right)D^{\frac{1}{2}}T
\end{aligned}
$$

The Covariance

$$
\begin{aligned}
Cov(x) &= \frac{1}{|2\pi|^{\frac{N}{2}}} T' D^{\frac{1}{2}} \left(\int_z zz' \exp(-\frac{1}{2} z' z) dz \right) D^{\frac{1}{2}} T \\
&= \frac{1}{|2\pi|^{\frac{N}{2}}} T' D^{\frac{1}{2}} \left(|2\pi|^{\frac{N}{2}} I \right) D^{\frac{1}{2}} T \\
&= T' D^{\frac{1}{2}} D^{\frac{1}{2}} T \\
&= T' D T \\
&= \Sigma
\end{aligned}
$$

Isodensity Contours

$$p(x) = \frac{1}{|2\pi|^{\frac{N}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu))$$

- μ mean vector
- Σ covariance matrix

The isodensity contours are defined by the sets $\{x \mid p(x) = constant\}$. If $p(x) = constant$, this implies that for some r,

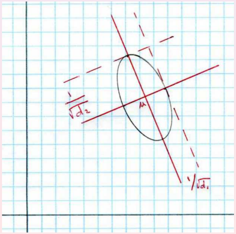
$$(x-\mu)'\Sigma^{-1}(x-\mu) = r^2$$

This is an equation of the surface of an ellipsoid with center μ.

The Ellipsoid

$$\begin{aligned} & \{x \mid (x-\mu)'\Sigma^{-1}(x-\mu) = r^2\} \\ = \; & \{x \mid (x-\mu)'T'D^{-1}T(x-\mu) = r^2\} \\ = \; & \{x \mid [D^{-\frac{1}{2}}T(x-\mu)]'[D^{-\frac{1}{2}}T(x-\mu)] = r^2\} \end{aligned}$$

If $d_1 > d_2$ and $D^{-1} = \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix}$



The Gaussian Classifier

The Gaussian Classifier assumes that the class conditional density functions are Multivariate Normal. It assigns a vector x to class c^1 when

$$\frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma_1|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)\right) P(c^1) \quad > \\ \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma_2|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)\right) P(c^2)$$

The Gaussian Classifier

Since *log* is monotonically increasing we can take the log on both sides of the inequality and maintain the inequality. Assign vector x to class c_1 when

$$-\frac{1}{2}(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1) - \frac{1}{2}\log|\Sigma_1| + \log P(c^1) \quad > \\ -\frac{1}{2}(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2) - \frac{1}{2}\log|\Sigma_2| + \log P(c^2)$$

Assign vector x to class c_1 when

$$(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1) + \log|\Sigma_1| - 2\log P(c^1) \quad < \\ (x-\mu_2)'\Sigma_2^{-1}(x-\mu_2) + \log|\Sigma_2| - 2\log P(c^2)$$

The Gaussian Classifier

When $|\Sigma_1| = |\Sigma_2|$ and $P(c^1) = P(c^2)$, then assign vector x to class c_1 when

$$(x - \mu_1)'\Sigma_1^{-1}(x - \mu_1) \quad < \quad (x - \mu_2)'\Sigma_2^{-1}(x - \mu_2)$$

The left hand side is the Mahalanobis distance between x and μ_1.
The right hand side is the Mahalanobis distance between x and μ_2.

The Gaussian Classifier

When $\Sigma_1 = \Sigma_2 = \Sigma$ and $P(c^1) = P(c^2)$, then assign vector x to class c_1 when

$$
\begin{aligned}
(x-\mu_1)'\Sigma^{-1}(x-\mu_1) &< (x-\mu_2)'\Sigma^{-1}(x-\mu_2) \\
x'\Sigma^{-1}x - 2\mu_1'\Sigma^{-1}x + \mu_1'\Sigma^{-1}\mu_1 &< x'\Sigma^{-1}x - 2\mu_2'\Sigma^{-1}x + \mu_2'\Sigma^{-1}\mu_2 \\
-2\mu_1'\Sigma^{-1}x + \mu_1'\Sigma^{-1}\mu_1 &< -2\mu_2'\Sigma^{-1}x + \mu_2'\Sigma^{-1}\mu_2 \\
2(\mu_2-\mu_1)'\Sigma^{-1}x &< \mu_2'\Sigma^{-1}\mu_2 - \mu_1'\Sigma^{-1}\mu_1 \\
(\mu_2-\mu_1)'\Sigma^{-1}x &< (\mu_2-\mu_1)'\Sigma^{-1}(\frac{\mu_1+\mu_2}{2}) \\
(\mu_2-\mu_1)'\Sigma^{-1}(x-\frac{\mu_1+\mu_2}{2}) &< 0
\end{aligned}
$$

Fisher Linear Discriminant

$$v = \Sigma_W^{-1}(\mu_1 - \mu_2)$$

Assign x to class 1 if

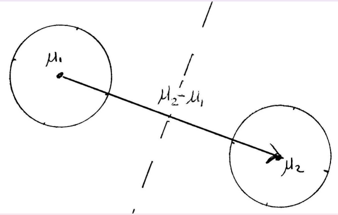
$$\begin{aligned} v'x &\geq \theta \\ \left(\Sigma_W^{-1}(\mu_1 - \mu_2)\right)' x &\geq \theta \\ (\mu_1 - \mu_2)'\Sigma_W^{-1\prime} x &\geq \theta \\ (\mu_1 - \mu_2)'\Sigma_W^{-1} x &\geq \theta \\ (\mu_2 - \mu_1)'\Sigma_W^{-1} x &< \theta \end{aligned}$$

When $\Sigma_1 = \Sigma_2$, the Gaussian classifier is a linear classifier and identical to the Fisher Linear Discriminant Classifier since $\Sigma_W = \Sigma_1 = \Sigma_2$

The Gaussian Classifier

When $\Sigma = I$, assign vector x to class c_1 when

$$(\mu_2 - \mu_1)'(x - \frac{\mu_1 + \mu_2}{2}) \quad < \quad 0$$



The dashed line represents the hyperplane passing through $\frac{\mu_1+\mu_2}{2}$ and perpendicular to its normal $\mu_2 - \mu_1$.
When x is to the left of the hyperplane, classify to class 1.
When x is to the right of the hyperplane, classify to class 2.

Pearson Distribution System

$$
\begin{aligned}
\frac{d}{dx}\log p(x) &= -\frac{x-a}{b_2x^2+b_1x+b_0} \\
\log p(x) &= -\int \frac{x-a}{b_2x^2+b_1x+b_0}dx + A \\
p(x) &= Ke^{-\int \frac{x-a}{b_2x^2+b_1x+b_0}dx}
\end{aligned}
$$

Assuming $b_2 \neq 0$, use partial fraction expansion

$$
p(x) = Ke^{-\int \frac{1}{b(a_1-a_2)}\left[\frac{a+a_1}{x-a_1}-\frac{a+a_2}{x-a_2}\right]dx}
$$

Pearson Type I

$a_1 < 0 < a_2$

$$\begin{aligned} p(x) &= K(x-a_1)^{m_1}(a_2-x)^{m_2}, \; a_1 \le x \le a_2 \\ m_1 &= \frac{a+a_1}{b(a_2-a_1)} \\ m_2 &= -\frac{a+a_2}{b(a_2-a_1)} \end{aligned}$$

Translating and scaling to put $0 \le x \le 1$

$$p(x) = \frac{1}{B(m,n)} x^{m-1}(1-x)^{n-1}$$

Pearson Type II

$a_1 < 0 < a_2$ and $m = m_1 = m_2$

$$p(x) = K(x - a_1)^m (a_2 - x)^m, \; a_1 \leq x \leq a_2$$

Translating and scaling to put $0 \leq x \leq 1$

$$p(x) = \frac{1}{B(m, m)} x^{m-1} (1 - x)^{m-1}$$

Pearson Type III

$b_2 = 0$ and $b_1 \neq 0$

$$
\begin{aligned}
\frac{d}{dx}\log p(x) &= -\frac{x-a}{b_1 x + b_0} \\
&= -\left[\frac{1}{b_1} - \frac{a + b_0/b_1}{b_0 + b_1 x}\right] \\
\log p(x) &= -\int \frac{1}{b_1} - \frac{a + b_0/b_1}{b_0 + b_1 x} dx + A \\
p(x) &= Ke^{-\int \frac{1}{b_1} - \frac{a+b_0/b_1}{b_0+b_1 x} dx} \\
&= K(b_0 + b_1 x)^m e^{-x/b_1},\ x \geq -\frac{b_0}{b_1},\ \text{when } b_1 > 0 \\
m &= b_1^{-1}(b_0 b_1^{-1} + a)
\end{aligned}
$$

Pearson Type III

Translating and scaling so that $x \geq 0$

$$p(x) = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x}, \ \lambda > 0, \ 0 \leq x$$

Pearson Type IV

Roots are complex

$$\begin{aligned}\frac{d}{dx}\log p(x) &= -\frac{x-a}{b_2x^2+b_1x+b_0}\\ &= -\frac{x-a}{c_0+b_2(x+c_1)^2}\end{aligned}$$

$$p(x) = K[c_0+b_2(x+c_1)^2]^{-(2b_2)^{-1}} e^{\frac{a+c_1}{\sqrt{b_2c_0}}\tan^{-1}\frac{x+c_1}{\sqrt{c_0/b_2}}}$$

Translating and scaling,

$$p(x) = K\left(1+\frac{x^2}{\sigma^2}\right)^{-m} e^{-\nu\tan^{-1}\frac{x}{\sigma}}$$

Pearson Type V

$b_2 x^2 + b_1 x + b_0$ is a perfect square

$$\begin{aligned}
\frac{d}{dx}\log p(x) &= -\frac{x-a}{b_2(x+c)^2} \\
p(x) &= K(x+c)^{-\frac{1}{b_2}} e^{\frac{a+c}{b_2(x+c)}}, \ x \geq -c
\end{aligned}$$

Pearson Type VI

Roots of $b_2x^2 + b_1x + b_0$ are real and negative; $a_1 < a_2 < 0$

$$\begin{aligned} p(x) &= K(x - a_1)^{m_1}(x - a_2)^{m_2},\ x > a_2 \\ m_1 &= < -1 \\ m_1 + m_2 &< 0 \end{aligned}$$

Pearson Type VII

$b_1 = a = 0,\ b_0, b_2 > 0$

$$
\begin{aligned}
\frac{d}{dx}\log p(x) &= -\frac{x}{b_2 x^2 + b_0} \\
p(x) &= K(b_0 + b_2 x^2)^{-\frac{1}{2b_2}}
\end{aligned}
$$

Definition

A function d is called a metric on a set X if and only if for every $x, y, z \in X$

- $d(x, y) \geq 0$
- $d(x, y) = 0$ if and only if $x = y$
- $d(x, y) = d(y, x)$
- $d(x, y) \leq d(x, z) + d(z, y)$

The space (X, d) is called a metric space.

Norm

Definition

The *norm* of a vector x is defined by

$$||x|| = \sqrt{x'x}$$

The Schwarz Inequality

Theorem

$$|a'b| \leq ||a|| \; ||b||$$

Proof:

$$\sum_{i=1}^{N} a_i^2 \sum_{j=1}^{N} b_j^2 - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (a_i b_j - b_i a_j)^2$$

$$= \sum_{i=1}^{N} a_i^2 \sum_{j=1}^{N} b_j^2 - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (a_i^2 b_j^2 - 2 a_i b_j b_i a_j + b_i^2 a_j^2)$$

Schwarz Inequality

$$
\begin{aligned}
&= \sum_{i=1}^{N} a_i^2 \sum_{j=1}^{N} b_j^2 - \frac{1}{2} \sum_{i=1}^{N} (a_i^2 \sum_{j=1}^{N} b_j^2 - 2 a_i b_i \sum_{j=1}^{N} a_j b_j + b_i^2 \sum_{j=1}^{N} a_j^2) \\
&= \sum_{i=1}^{N} a_i^2 \sum_{j=1}^{N} b_j^2 - \frac{1}{2} \left(\sum_{i=1}^{N} a_i^2 \sum_{j=1}^{N} b_j^2 - 2 \sum_{i=1}^{N} a_i b_i \sum_{j=1}^{N} a_j b_j + \right. \\
&\qquad \left. \sum_{i=1}^{N} b_i^2 \sum_{j=1}^{N} a_j^2 \right) \\
&= \left(\sum_{i=1}^{N} a_i b_i \right)^2
\end{aligned}
$$

Schwarz Inequality

$$\left(\sum_{i=1}^{N} a_i b_i\right)^2 = \sum_{i=1}^{N} a_i^2 \sum_{j=1}^{N} b_j^2 - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(a_i b_j - b_i a_j)^2$$

$$\leq \sum_{i=1}^{N} a_i^2 \sum_{j=1}^{N} b_j^2$$

$$(a'b)^2 \leq ||a||^2 \; ||b||^2$$

Norm

Theorem

$$||a+b|| \leq ||a|| + ||b||$$

Proof:

$$\begin{aligned} ||a+b||^2 &= (a+b)'(a+b) = a'a + 2a'b + b'b \\ &= ||a||^2 + 2a'b + ||b||^2 \\ &\leq ||a||^2 + 2|a'b| + ||b||^2 \\ &\leq ||a||^2 + 2||a||\;||b|| + ||b||^2 \\ &\leq (||a|| + ||b||)^2 \end{aligned}$$

Theorem

The norm of a vector difference is a metric. Define $d(x, y) = ||x - y||$. Then d is a metric.

$$\begin{aligned}
d(x, y) &= ||x - y|| = \sqrt{(x - y)'(x - y)} \geq 0 \\
0 &= ||x - y|| \text{ if and only if } x - y = 0 \\
d(x, y) &= ||x - y|| = \sqrt{(x - y)'(x - y)} \\
&= \sqrt{(y - x)'(y - x)} = d(y, x) \\
d(x, y) &= ||x - y|| = ||(x - z) + (z - y)|| \\
&\leq ||x - z|| + ||z - y|| = d(x, z) + d(z, y)
\end{aligned}$$

Theorem

Let A be a positive definite symmetric matrix. Define $\rho(x, y) = \sqrt{(x-y)' A (x-y)}$, then ρ is a metric.

Proof: Let $A = T'DT$. Then

$$
\begin{aligned}
\rho^2(x, y) &= (x-y)'A(x-y) = (x-y)'T'DT(x-y) \\
&= (x-y)'T'D^{\frac{1}{2}}D^{\frac{1}{2}}T(x-y) = \\
&= (D^{\frac{1}{2}}T(x-y))'(D^{\frac{1}{2}}T(x-y)) \\
&= \left((D^{\frac{1}{2}}Tx) - (D^{\frac{1}{2}}Ty)\right)' \left((D^{\frac{1}{2}}Tx) - (D^{\frac{1}{2}}Ty)\right) \\
&= d^2(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Ty)
\end{aligned}
$$

$$\rho(x, y) = d(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Ty)$$

(1)
Since $d(u, v) \geq 0$ for every u, v,
$d(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Ty) \geq 0$ so that $\rho(x, y) \geq 0$.

$\rho(x, y) = 0$ if and only if $d(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Ty) = 0$.
And $d(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Ty) = 0$ if and only if $D^{\frac{1}{2}}Tx = D^{\frac{1}{2}}Ty$.
$D^{\frac{1}{2}}Tx = D^{\frac{1}{2}}Ty$ if and only if $x = y$.

Quadratic Forms and Metrics

$$\rho(x,y) = d(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Ty)$$

(2)

$$\begin{aligned}\rho(x,y) &= d(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Ty)\\ &= d(D^{\frac{1}{2}}Ty, D^{\frac{1}{2}}Tx)\\ &= \rho(y,x)\end{aligned}$$

Quadratic Forms and Metrics

$$\rho(x, y) = d(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Ty)$$

(3)

$$\begin{aligned} \rho(x, y) &= d(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Ty) \\ &\leq d(D^{\frac{1}{2}}Tx, D^{\frac{1}{2}}Tz) + d(D^{\frac{1}{2}}Tz, D^{\frac{1}{2}}Ty) \\ &\leq \rho(x, z) + \rho(z, y) \end{aligned}$$

Multivariate Normal

$$\begin{aligned} p(x) &= \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)) \\ &= \frac{1}{|2\pi|^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}\rho(x,\mu;\Sigma^{-1})^2 \end{aligned}$$

Notice that $\exp(-\frac{1}{2}u^2)$ is a monotonically decreasing function of u. The Multivariate Normal density is just one density that converts a distance to a density through a monotonically decreasing function.

General Ellipsoidally Symmetric Multivariate Forms

Let f be any non-negative decreasing function in its tail and A a symmetric positive definite matrix. Then

$$p(x) = Kf\left(\sqrt{(x-\mu)'A(x-\mu)}\right)$$

with an appropriate value for K is a multivariate density function.

General Ellipsoidally Symmetric Multivariate Forms

$f : [0, \infty] \to [0, \infty]$ monotonically decreasing in its tail with

$$\int_r r^{N+1} f(r) dr \quad < \quad \infty$$

Define

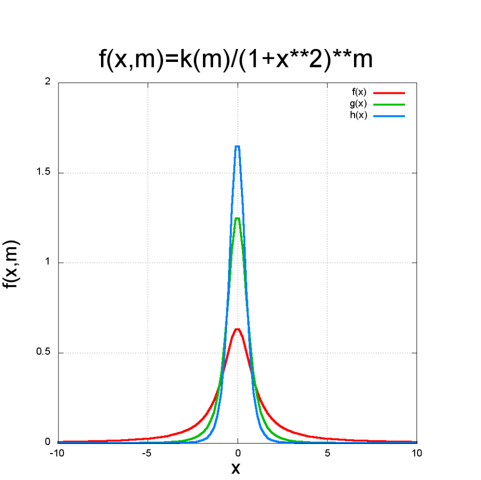
$$p(x) \quad = \quad Kf\left(\sqrt{(x-\mu)' A (x-\mu)}\right)$$

If $\int p(x) dx = 1$, then

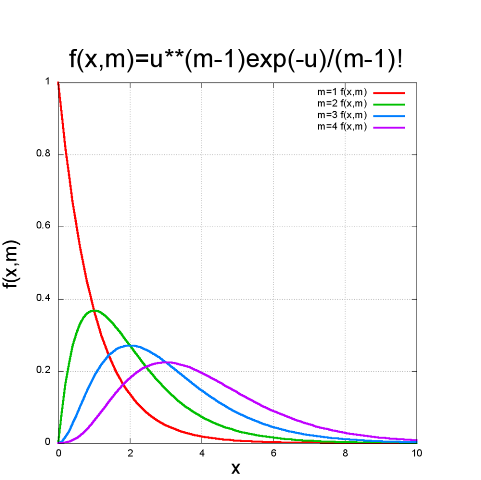
$$\begin{aligned} K &= \frac{\Gamma(N/2)}{2\pi^{N/2} \int_r r^{N-1} f(r) dr} |A|^{\frac{1}{2}} \\ A &= \frac{\int_r r^{N+1} f(r) dr}{N \int_r r^{N-1} f(r) dr} \Sigma^{-1} \end{aligned}$$

f	K	A
e^{-u^2}	$\frac{1}{(2\pi)^{N/2}}\|A\|^{\frac{1}{2}}$	Σ^{-1}
$(1+u^2)^{-m}$ $2m-N-2>0$	$\frac{\Gamma(m)}{\pi^{N/2}\Gamma(m-N/2)}\|A\|^{\frac{1}{2}}$	$\frac{1}{2m-N-2}\Sigma^{-1}$
$u^{m-1}e^{-u}$ $N+m-1>0$	$\frac{\Gamma(N/2)}{2\pi^{N/2}\Gamma(N+m-1)}\|A\|^{\frac{1}{2}}$	$\frac{(N+m-1)(N+m-2)}{N}\Sigma^{-1}$

Functional Forms



Gamma Distribution



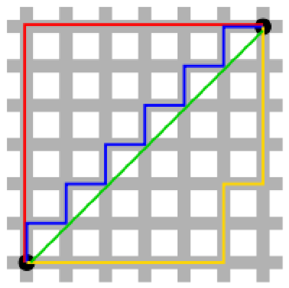
Metrics

$$
\begin{aligned}
\rho_{Manhattan}(x, y) &= \sum_{n=1}^{N} |x_n - y_n| \\
\rho_{Euclidean}(x, y) &= ||x - y||^2 \\
\rho_{Chebyshev}(x, y) &= \max_{n=1,\ldots,N} |x_n - y_n|
\end{aligned}
$$

Manhattan Metric

$$\rho_{Manhattan}(x, y) = \sum_{n=1}^{N} |x_n - y_n|$$

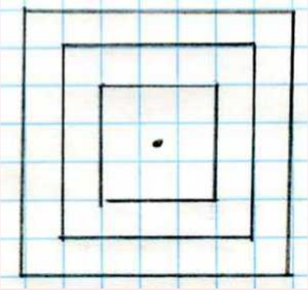
Also called the taxi cab metric and the city block metric.



Chebyshev

$$\rho_{Chebyshev}(x, y) = \max_{n=1,\ldots,N} |x_n - y_n|$$

Also called the chess board distance.



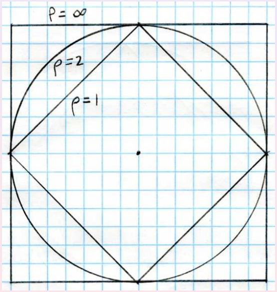
Minkowski Distance

$$\rho_{Minkowski}(x, y) = \left(\sum_{n=1}^{N} |x_n - y_n|^p \right)^{\frac{1}{p}}$$

- $p = 1$ Manhattan
- $p = 2$ Euclidean
- $p \to \infty$ Chebyshev

Minkowski Distance

$$\rho_{Minkowski}(x, y) = \left(\sum_{n=1}^{N} |x_n - y_n|^p \right)^{\frac{1}{p}}$$



Minkowski Distance

$$\rho_{Minkowski}(x, y) = \left(\sum_{n=1}^{N} |x_n - y_n|^p \right)^{\frac{1}{p}}$$

Let $|x_k - y_k| > |x_n - y_n|, \; n = 1, \ldots, N$ and $n \neq k$

$$\begin{aligned} \lim_{p \to \infty} \left(\sum_{n=1}^{N} |x_n - y_n|^p \right)^{\frac{1}{p}} &= |x_k - y_k| \lim_{p \to \infty} \left[\sum_{n=1}^{N} \left(\frac{|x_n - y_n|}{|x_k - y_k|} \right)^p \right]^{\frac{1}{p}} \\ &= |x_k - y_x| \end{aligned}$$

General Parametric Probability Densities

Let f_k be a strictly monotonically decreasing function whose (N+2) order moment exists. Define

$$p(x \mid c^k) = m_k f_k(\rho_k(x, \mu_k))$$

Assign a vector x to class c_1 when

$$m_1 f_1(\rho_1(x, \mu_1))P(c^1) > m_2 f_2(\rho_2(x, \mu_2))P(c^2)$$

Using Only Distance Functions

Consider

$$\begin{pmatrix} \rho_1(x, \mu_1) \\ \rho_2(x, \mu_2) \end{pmatrix}$$

to be a feature vector.

Training Data: 2D Features

$$< x_1, \ldots, x_M > \text{ class 1 } x_m \in \mathbb{R}^J$$
$$< y_1, \ldots, y_N > \text{ class 2 } y_n \in \mathbb{R}^J$$

Form the class 1 feature vectors

$$\begin{pmatrix} \rho_1(x_1, \mu_1) \\ \rho_2(x_1, \mu_2) \end{pmatrix}, \ldots, \begin{pmatrix} \rho_1(x_M, \mu_1) \\ \rho_2(x_M, \mu_2) \end{pmatrix}$$

Form the class 2 feature vectors

$$\begin{pmatrix} \rho_1(y_1, \mu_1) \\ \rho_2(y_1, \mu_2) \end{pmatrix}, \ldots, \begin{pmatrix} \rho_1(y_N, \mu_1) \\ \rho_2(y_N, \mu_2) \end{pmatrix}$$

Quantize

Let q_i be a quantizer for feature i.

Class 1 feature vectors

$$\begin{pmatrix} q_1(\rho_1(x_1, \mu_1)) \\ q_2(\rho_2(x_1, \mu_2)) \end{pmatrix}, \ldots, \begin{pmatrix} q_1(\rho_1(x_M, \mu_1)) \\ q_2(\rho_2(x_M, \mu_2)) \end{pmatrix}$$

Class 2 feature vectors

$$\begin{pmatrix} q_1(\rho_1(y_1, \mu_1)) \\ q_2(\rho_2(y_1, \mu_2)) \end{pmatrix}, \ldots, \begin{pmatrix} q_1(\rho_1(y_N, \mu_1)) \\ q_2(\rho_2(y_N, \mu_2)) \end{pmatrix}$$

Use discrete probability Bayesian classification methodology

Training Data: 1D Features

$$< x_1, \ldots, x_M > \text{class 1}$$
$$< y_1, \ldots, y_N > \text{class 2}$$

Form the class 1 feature values

$$(\rho_1(x_1, \mu_1)), \ldots, (\rho_1(x_M, \mu_1))$$

Form the class 2 feature values

$$(\rho_2(y_1, \mu_2)), \ldots, (\rho_2(y_N, \mu_2))$$

Quantize

Same quantizer q for both classes

Form the class 1 feature values

$$(q(\rho_1(x_1,\mu_1))),\ldots,(q(\rho_1(x_M,\mu_1)))$$

Form the class 2 feature values

$$(q(\rho_2(y_1,\mu_2))),\ldots,(q(\rho_2(y_N,\mu_2)))$$

Non-parametric Probability Model

- Quantize the feature values for class 1 and class 2
- Construct the histogram h_1 for the class 1 feature values
- Construct the histogram h_2 for the class 2 feature values
- Normalize histograms so that they each sum to 1
- $P(c^1, x) = h_1(q(\rho(x, \mu_1)))P(c^1)$
  $P(c^2, x) = h_2(q(\rho(x, \mu_2)))P(c^2)$
- Use discrete probability Bayesian classification

The Euclidean Distance Geometry

- In 2-D, $\{x \mid q(\rho(x, \mu)) = k\}$ is a ring around μ
- In N-D, $\{x \mid q(\rho(x, \mu)) = k\}$ is a spherical shell around μ
  - That spherical shell associated with k has a probability $P(q(\rho(x, \mu) = k)$ There are many patterns the probability can cover
    - When k is small the probability that a point x falls in the spherical shell can be large
    - As k increases the probability can grow small
    - Like the Mahalanobis distance of x to μ
    - But after a while of growing smaller with increasing k it can grow larger and then smaller
    - Probability can be large near μ or near a spherical shell associated with a k of intermediate distance to μ

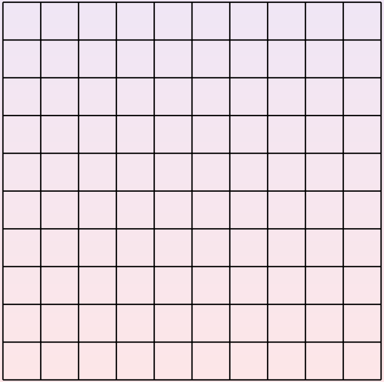
The Euclidean Distance Geometry

- Let h be the normalized histogram obtained from the training data
- $h(k)$ is the probability of an x falling into the spherical shell associated with k
- $P(q(\rho(x, \mu) = k) = h(k)$ says that the probability of an x falling into the spherical shell associated with k is $h(k)$
- When h is computed from the training set its values are governed from the training set and can be arbitrary not following any pre-given pattern

Viewing the Quantized Space

We consider looking the quantized distance to spherical shells for class 1 and quantized distance to spherical shells for class 2

- q is equal interval quantizing
- x-axis is $q(\rho(x, \mu_1))$
- y-axis is $q(\rho(x, \mu_2))$



Euclidean Distance: Spherical Shells

- $q(\rho(x, \mu_1)) = i$
  - x is in the i^{th} spherical shell for μ_1
- $q(\rho(x, \mu_2)) = j$
  - x is in the j^{th} spherical shell for μ_2
- This does not state that the spherical shells intersect
- The spherical shells for the quantized range may never intersect

Max Distance: Hyperbox Shells

- $q(\rho(x, \mu_1)) = i$
  - x is in the i^{th} hyperbox shell for μ_1
- $q(\rho(x, \mu_2)) = j$
  - x is in the j^{th} hyperbox shell for μ_2
- This does not state that the hyperbox shells intersect
- The hyperbox shells for the quantized range may never intersect

Quantized Class Conditional Probabilities

- $q(P(x \mid c_1)) = i$
  - $P(x \mid c_1)$ is in the i^{th} quantized probability interval for given class c_2
- $q(P(x \mid c_2)) = j$
  - $P(x \mid c_2)$ is in the j^{th} quantized probability interval for given class c_2