MULTIVARIATE NORMAL

Univariate Normal (distribution of a random variable)

- Each data point is a scalar
- Formally, given some data $\{x_i\}_{i=1}^{n} \in \mathbb{R}$

random variable $\leftarrow x \sim N(\mu, \sigma^2)$

Mean

$$\mu = E(x)$$
$$= \begin{cases} \frac{1}{n}\sum_{i=1}^{n} x_i & \langle \text{discrete} \rangle \\ \int x_i \cdot p(x_i)\, dx & \langle \text{continuous} \rangle \end{cases}$$

Variance

$$\sigma^2 = E(x-\mu)^2$$
$$= \begin{cases} \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 & \langle \text{discrete} \rangle \\ \int (x_i-\mu)^2\, p(x_i)\, dx & \langle \text{continuous} \rangle \end{cases}$$

Multivariate Normal (joint distribution of many random variables)

- Each data point is a vector
- Formally, given some data $\{x_i\}_{i=1}^{n} \in \mathbb{R}^d$

vector of random variables $\leftarrow x \sim N(\mu, \Sigma)$

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}$$

Mean

$$\mu = E(x) = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_d) \end{bmatrix} = \begin{cases} \begin{bmatrix} \frac{1}{m}\sum^{m} x_1 \\ \vdots \\ \frac{1}{m}\sum x_d \end{bmatrix} & \langle \text{discrete} \rangle \\ \begin{bmatrix} \int x_1^{(i)}\, p(x_1^{(i)})\, dx_1 \\ \vdots \\ \int x_d^{(i)}\, p(x_d^{(i)})\, dx_d \end{bmatrix} & \langle \text{continuous} \rangle \end{cases}$$

Covariance matrix

$$\Sigma = E\left((x-\mu)(x-\mu)^T\right) = \begin{bmatrix} \lambda_{11} & \cdots & \lambda_{1d} \\ \vdots & \ddots & \vdots \\ \lambda_{d1} & \cdots & \lambda_{dd} \end{bmatrix}$$

$cov(x_1, x_d)$: $E[(x_1 - \mu_1)(x_d - \mu_d)]$

variance of x_i: $E(x_i - \mu_1)^2$

PROPERTIES OF MULTIVARIATE NORMAL

1. Independent Normal Distributions

- The covariance matrix is diagonal
- Formally, given some data $\{x_i\}_{i=1}^{n} \in \mathbb{R}^d$, where $x_i \perp x_j \quad \forall\, i \neq j$ (independent)

$$x \sim N(\mu, \Sigma)$$

Diagonal Covariance Matrix

$$\Sigma = E\left[(x-\mu)(x-\mu)^T\right] = \begin{bmatrix} \lambda_{11} & & 0 \\ & \ddots & \\ 0 & & \lambda_{dd} \end{bmatrix}$$

Proof $Cov(X,Y) = 0$ if X independent Y:

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$
$$= E(X)E(Y) - E(X)E(Y)$$

2. Marginal Distributions

- The marginal distributions that are components of multivariate normal distribution is also normal.
- Formally, we can partition vector and matrix of:

$$x \sim N(\mu, \Sigma)$$

vector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\quad \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ matrix

x_1 is also Gaussian

$$x_1 \sim N(\mu_1, \Sigma_{11})$$

3. Conditional Distributions

- The conditional distributions of any components of multivariate normal distribution is also normal
- Formally, conditional on first component of vector $x \sim N(\mu, \Sigma)$:

$$x_1 \mid x_2 = a \sim N(\mu_{1|2}, \Sigma_{1|2})$$

Conditional Mean

$$\mu_{1|2} = E(x_1 \mid x_2 = a)$$
$$= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(a - \mu_2)$$

Conditional Covariance matrix

$$\Sigma_{1|2} = cov(x_1 \mid x_2 = a)$$
$$= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$\Sigma_{1|2}$ is Schur complement of Σ_{22} in Σ

4. Linear Transform

- Linear transform multivariate normal distribution give a another multivariate normal
- Formally, given some data $\{x_i\}_{i=1}^{n} \in \mathbb{R}^d$:

$$x \sim N(\mu, \Sigma)$$

$a^T (x-\mu)(x-y)^T a$

and : $z = Ax + b$

then : $z \sim N(\mu_z, \Sigma_z)$

Transformed Mean

$$\begin{aligned} \mu_z &= E(z) \\ &= E(Ax + b) \\ &= A . E(x) + b \\ &= A\mu + b \end{aligned}$$

Transformed Covariance Matrix

$$\begin{aligned} \Sigma_z &= E\left[(z-\mu_z)(z-\mu_z)^T\right] \\ &= E\left[A(x-\mu_z)\left[A(x-\mu_z)\right]^T\right] \\ &= E\left[A(x-\mu_z)(x-\mu_z)^T A^T\right] \\ &= A \Sigma A^T \end{aligned}$$

5. PCA under the lense of multivariate normal distributions

- Recall in PCA, we try to find principal components with maximum variance.

  i) Given some data $\{x_i\}_{i=1}^{n} \in \mathbb{R}^d$, assume the data is normally distributed and centralized :

$$x \sim N(0, \Sigma)$$

  ii) Eigendecompose covariance matrix Σ :

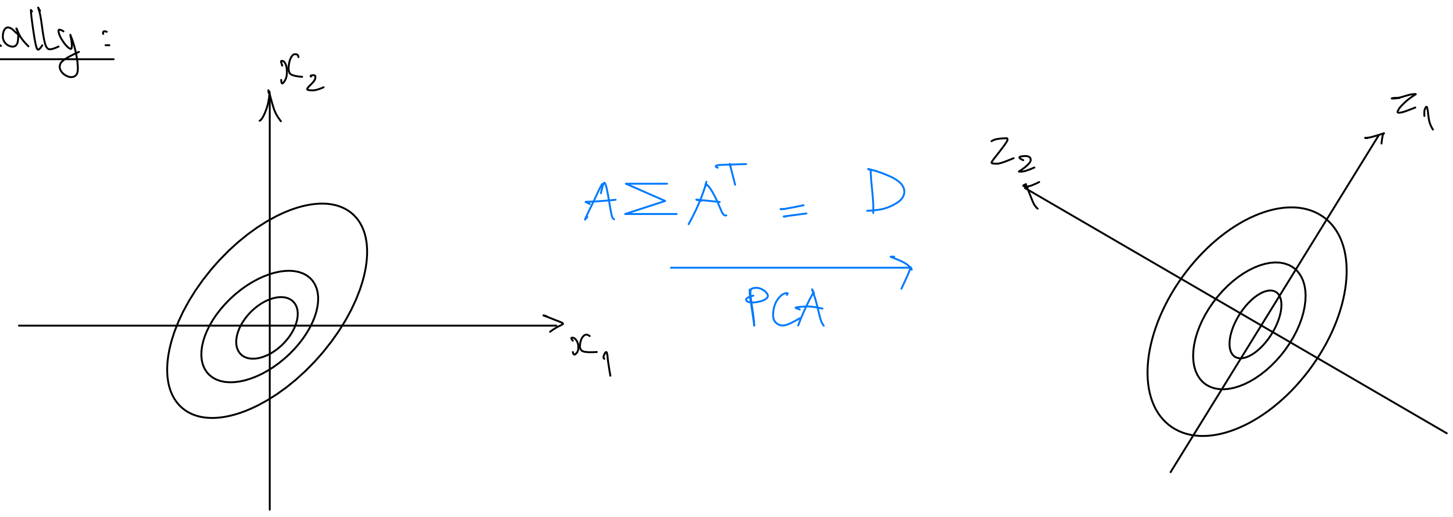
$$\Sigma = Q D Q^T$$

  iii) Let say $A = Q^T$ and $z = Ax + b$, then :

$$z \sim N(0, D)$$

new coordinate system

- Visually :



MULTIVARIATE NORMAL (NATURAL FORM)

Standard form: $x \sim N(\mu, \Sigma)$

$$p(x) = \frac{1}{Z}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$(2\pi)^{d/2}\det(\Sigma)^{1/2}$

Natural form: $x \sim \bar{N}(b, Q)$

$$p(x) = \frac{1}{C}\exp\left(-\frac{1}{2}x^T Q x + b^T x\right)$$

$$\propto \exp\left(-\frac{1}{2}x^T Q x + b^T x\right)$$

$Q = \Sigma^{-1}$ (precision matrix) $\qquad b = \Sigma^{-1}\mu$

$C = \exp\left(-\frac{1}{2}b^T Q^{-1} b\right) / Z$

(normalization constant)

Proof Standard form equals Natural form

$$p(x) = \frac{1}{Z}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$$\propto \exp\left(-\frac{1}{2}\left(x^T\Sigma^{-1}x - 2x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu\right)\right)$$

$$\propto \exp\left(-\frac{1}{2}x^T\underbrace{\Sigma^{-1}}_{Q}x + x^T\underbrace{\Sigma^{-1}\mu}_{b}\right)$$

Why use natural form?

- Easier to derive conditional distributions
- For example, given some data $\{x_i\}_{i=1}^{n} \sim N(\mu, \Sigma)$.

Then the marginal distribution is given by:

i) In Standard Form:

$$x_1 \sim N(\mu_1 \mid \underbrace{\Sigma_{11}}_{Cov(x_1)})$$

And the conditional distribution is given by:

i) In Standard Form:

$$x_{1|2=a} \sim N(\mu_{1|2}, \Sigma_{1|2})$$

- $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$
- $\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

$Cov(x_1 | x_2) = Q_{11}^{-1}$

ii) In Natural Form:

$$x_{1|2=a} \sim \bar{N}(b_1 - Q_{12}x_2, Q_{11})$$

Proof that $x_1 \mid x_2 = a \sim \overline{N}(b_1 - Q_{12} x_2 \,,\, Q_{11})$

Let $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Then in Natural Form:

- $p(x) \propto \exp\left[-\frac{1}{2}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}^T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right]$

$$\propto \exp\left[-\frac{1}{2}\left(x_1^T Q_{11} x_1 + x_1^T Q_{12} x_2 + x_2^T Q_{21} x_1 + x_2^T Q_{22} x_2\right) + \left(b_1^T x_1 + b_2^T x_2\right)\right]$$

$Q_{12} = Q_{21}$

$$\propto \exp\left[-\frac{1}{2}\left(x_1^T Q_{11} x_1 + 2 x_1^T Q_{12} x_1 + x_2^T Q_{22} x_2\right) + \left(b_1^T x_1 + b_2^T x_2\right)\right]$$

Conditioned on x_2 (remove x_2 out of the equation)

- $p(x_1 \mid x_2) \propto \exp\left[-\frac{1}{2}\left(x_1^T Q_{11} x_1 + 2 x_1^T Q_{12} x_1\right) + b_1^T x_1\right]$

$$\propto \exp\left[-\frac{1}{2} x_1^T Q_{11} x_1 + x_1^T \left(b_1 - Q_{12} x_1\right)\right]$$

Therefore, $x_1 \mid x_2 = a \sim \overline{N}(b_1 - Q_{12} x_1 \,,\, Q_{11})$

What do Standard Form and Natural Form tell us about the relationship between different features

- Covariance matrix $\Sigma = [\sigma_{ij}]_{ij}$ measures marginal independence

  $\sigma_{ij} = 0$ if $x_i \perp x_j$

- Precision matrix $Q = \Sigma^{-1} = [q_{ij}]_{ij}$ measures conditional independence

  $q_{ij} = 0$ if $x_i \perp x_j \mid x_{\neg ij}$ not i and j

What this means in term of probability?

- Marginal Independence: $x_1 \perp x_2 \Leftrightarrow P(x_1, x_2) = P(x_1) \cdot P(x_2)$
- Conditional Independence: $x_1 \perp x_2 \mid x_3 \Leftrightarrow P(x_1, x_2 \mid x_3) = P(x_1 \mid x_3) \cdot P(x_2 \mid x_3)$

Proof $x_1 \perp x_2 \mid x_3 \Leftrightarrow P([x_1, x_2] \mid x_3) = P(x_1 \mid x_3) \cdot P(x_2 \mid x_3)$

- Partition $Q = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} \\ \vdots & \ddots & \vdots \\ Q_{31} & \cdots & Q_{33} \end{bmatrix}$ $\quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$

- Using previous work, we can prove that:

$$[x_1 \ x_2] \mid x_3 \sim \overline{N}\Big(\underbrace{\boxed{b_{1:2} - Q_{1:2,3}\, x_3}}_{\tilde{b}} \ , \ \underbrace{\boxed{Q_{1:2,\,1:2}}}_{\tilde{Q}}\Big)$$

- Now we analyze $p([x_1, x_2] \mid x_3)$:

$$p([x_1, x_2] \mid x_3) \propto \exp\left[-\frac{1}{2}\left(x_1^T Q_{11} x_1 + 2\, x_1^T Q_{12}\, x_2 + x_2^T Q_{22}\, x_2\right) + \left(\tilde{b}_1^T x_1 + \tilde{b}_2^T x_2\right)\right]$$

$$\propto \underbrace{\exp\left(-\frac{1}{2} x_1^T Q_{11}\, x_1 + \tilde{b}_1^T x_1\right)} \cdot \underbrace{\exp\left(-\frac{1}{2} x_2^T Q_{22}\, x_2 + \tilde{b}_2^T x_2\right)}$$

$$\propto \quad p(x_1 \mid x_3) \quad \cdot \quad p(x_2 \mid x_3)$$

So $x_1 \perp x_2 \mid x_3$