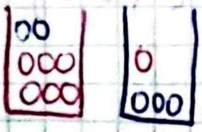
Principio de probabilidad.

1. Para $P(B=r)=0.4$ y $P(B=b)=0.6$

$$P(B=r)+P(B=b)=1$$

Según figura encontrar:

Nota: Tener en cuenta regla del producto.



a: apple
o: orange
B: Box.
f: fruit.

Probabilidades a encontrar (Marginal):

A. $P(f=a)=P(f=a, B=b)+P(f=a, B=r)$

Regla producto $= P(f=a \mid B=b)\cdot P(B=b)+P(f=a \mid B=r)\cdot P(B=r)$

$$=\frac{3}{4}\times 0.6+\frac{2}{8}\times 0.4$$

$$\approx 0.55$$

B. $P(f=o)=P(f=o \mid B=b)\cdot P(B=b)+P(f=o \mid B=r)\cdot P(B=r)$

$$=\frac{1}{4}\times 0.6+\frac{6}{8}\times 0.4 \approx 0.45$$

C. $P(B=r \mid F=o)$:
Aplicando Bayes.

$$P(B=r \mid f=o)=\frac{P(f=o \mid B=r)\cdot P(B=r)}{P(f=o)}$$

$$=\frac{6/8\times 0.4}{0.45}\approx 0.67$$

D. $P(B=b \mid f=o)=\frac{1/4\times 0.6}{0.45}\approx 0.33$

E. $P(B=r \mid f=a)=\frac{2/8\times 0.4}{0.55}\approx 0.18$

f. $P(B=b \mid f=a) = \dfrac{3/4 \times 0.6}{0.55} \approx 0.82.$

2. Demuestre que:

Nota: $\mathcal{E}\{x\} = \mu_x$

$\mathcal{E}\{\mu_x^2\} = \mu_x^2$

$\mathcal{E}\{cte\} = cte.$

$$\mathrm{Var}\{x\} = \mathcal{E}\{x^2\} - (\mathcal{E}\{x\})^2$$

$$= \mathcal{E}\{(x-\mu_x)^2\}$$

$$= \mathcal{E}\{x^2 - 2\mu_x x + \mu_x^2\}$$

$$= \mathcal{E}\{x^2\} - 2\mu_x \mathcal{E}\{x\} + \mathcal{E}\{\mu_x^2\}$$

$$= \mathcal{E}\{x^2\} - 2\mathcal{E}\{x\}\mathcal{E}\{x\} + \mu_x^2$$

$$= \mathcal{E}\{x^2\} - 2\mathcal{E}^2\{x\} + \mathcal{E}^2\{x\}$$

$$= \mathcal{E}\{x^2\} - \mathcal{E}^2\{x\}$$

$$\mathrm{Cov}\{x,y\} = \mathcal{E}_{x,y}\{xy\} - \mathcal{E}\{x\}\mathcal{E}\{y\}$$

$$= \mathcal{E}_{x,y}\{(x-\mu_x)(y-\mu_y)\}$$

$$= \mathcal{E}_{x,y}\{xy - \mu_x y - x\mu_y + \mu_x\mu_y\}$$

$$= \mathcal{E}_{x,y}\{xy\} - \mu_x\mathcal{E}\{y\} - \mathcal{E}\{x\}\mu_y + \mu_x\mu_y$$

$$= \mathcal{E}_{x,y}\{xy\} - \mathcal{E}\{x\}\mathcal{E}\{y\} - \cancel{\mathcal{E}\{x\}\mathcal{E}\{y\}} + \cancel{\mathcal{E}\{x\}\mathcal{E}\{y\}}$$

$$= \mathcal{E}_{x,y}\{xy\} - \mathcal{E}\{x\}\mathcal{E}\{y\}$$

$$\mathrm{Cov}(\mathbf{x},\mathbf{y}) = \mathcal{E}_{xy}\{\mathbf{x}\mathbf{y}^T\} - \mathcal{E}\{\mathbf{x}\}\mathcal{E}\{\mathbf{y}^T\}$$

donde $\mathbf{x}$, $\mathbf{y}$ vectores aleatorios.

$\mathbf{x}, \mathbf{y} \in \mathbb{R}^{D \times 1}$.

$$= \mathcal{E}_{xy}\{(x-\mu_x)(y-\mu_y)^T\}$$

$$= \mathcal{E}_{xy}\{xy^T - \mu_x y^T - x\mu_y^T + \mu_x\mu_y^T\}$$

$$= \mathcal{E}_{xy}\{xy^T\} - \mu_x\mathcal{E}_{xy}\{y^T\} - \mathcal{E}_{x,y}\{x\}\mu_y^T + \mu_x\mu_y^T$$

$$= \mathcal{E}_{xy}\{xy^T\} - \mathcal{E}\{x\}\mathcal{E}\{y^T\} - \mathcal{E}\{x\}\mathcal{E}\{y^T\} + \mathcal{E}\{x\}\mathcal{E}\{y^T\}$$

$$= \mathcal{E}_{xy}\{xy^T\} - \mathcal{E}\{x\}\mathcal{E}\{y^T\}.$$

3. $\mathcal{E}\{x^2\} = \int_{-\infty}^{\infty} N(x|\mu,\sigma^2)\, x^2 dx = \mu^2+\sigma^2$

$$= \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Cambio variable $z = \frac{x-\mu}{\sigma}$;

$$= \int_{-\infty}^{\infty} (\sigma z+\mu)^2 \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz$$

$$= \sigma^2 \underbrace{\int_{-\infty}^{\infty} z^2 \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz}_{\sigma^2} + \underbrace{2\sigma\mu \int_{-\infty}^{\infty} z \cdot \frac{1}{2\pi} \exp\left(\frac{-z^2}{2}\right) dz}_{0}$$

$$+ \underbrace{\mu^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dz.}_{\mu^2}$$

$$= \sigma^2 \times 1 + 0 + \mu^2 \times 1 = \mu^2+\sigma^2$$

$$= \mu^2+\sigma^2.$$

$Var\{x\} = \sigma^2.$

$Var\{x\} = \mathcal{E}\{x^2\} - \{\mathcal{E}[x]\}^2$

$\mathcal{E}\{x\} = \mu$ - - - - - - - - - . Para una distribución normal.

$\mathcal{E}\{x^2\} = \mu^2+\sigma^2$ - - - - PD anterior.

fórmula varianza

$Var\{x\} = (\mu^2+\sigma^2) - \mu^2.$

$Var\{x\} = \sigma^2.$

1. Demostrar

$$\mathcal{E}\{x\} = \int_{-\infty}^{\infty} N(x|\mu, \sigma^2)\, x\, dx = \mu$$

$$\mathcal{E}\{x^2\} = \int_{-\infty}^{\infty} N(x|\mu, \sigma^2)\, x^2\, dx = \mu^2 + \sigma^2$$

$$\mathrm{Var}\{x\} = \sigma^2$$

1. $$\mathcal{E}\{x\} = \int_{-\infty}^{\infty} N(x|\mu, \sigma^2)\, x\, dx = \mu$$

Escribir función densidad probabilidad Normal.

$$N(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Sustituir en la integral del valor esperado.

$$\mathcal{E}\{x\} = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

Cambio variable.

$u = \frac{x-\mu}{\sigma}$, de modo que $du = \frac{dx}{\sigma}$ y $x = u\sigma + \mu$.

$$\mathcal{E}\{x\} = \int_{-\infty}^{\infty} (u\sigma + \mu) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

$$\mathcal{E}\{x\} = \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du + \sigma \int_{-\infty}^{\infty} u \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

1. $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = 1$, Entonces, $\mu \times 1 = 1$.

2. $\int_{-\infty}^{\infty} u \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = 0$

$$\mathcal{E}\{x\} = \mu \times 1 + \sigma \times 0 = \mu$$

4. Encontrar estimadores de máxima verosimilitud μ_{ML} y σ^2_{ML} utilizando log-verosimilitud para una muestra $x_1, x_2, \cdots, x_n$ de una distribución $N(\mu, \sigma^2)$.

Función log-verosimilitud.

$$\log L(\mu, \sigma^2) = \log\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right)$$

$$= \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right)$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

No depende μ.

Derivar respecto μ y σ^2 para maximizar log-verosimilitud.

Derivada respecto μ:

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial \log L}{\partial \mu} = 0 \Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\mu_{ML} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Derivada respecto σ^2:

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial \log L}{\partial \sigma^2} = 0 \Rightarrow \sigma^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_{ML})^2$$

$$\sigma^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_{ML})^2.$$

5. Sea $x_n = A + w_n$, con $A \in \mathbb{R}$ y $w_n \sim N(w_n | M, \sigma^2)$

Sea $\hat{A} = \frac{1}{N} \sum_{n=1}^{N} x_n$ Estimador 1.

$\tilde{A} = x_1$ Estimador 2.

¿Cuál es mejor estimador desde $\{x_1, x_2, \dots, x_n\}$?

Sesgo $\hat{A}_1 = b(\hat{A}) = \varepsilon\{\hat{A}\} - A$

$$= \varepsilon\left[\frac{1}{N}\sum_{n=1}^{N} x_n\right] - A = \frac{1}{N}\sum_{n=1}^{N} \varepsilon\{x_n\} - A.$$

$\varepsilon[x_n] = A$ $(x_n = A + w_n$ y $\varepsilon\{w_n\} = 0)$

$$= \frac{1}{N}\sum_{n=1}^{N} A - A = 0$$

$\hat{A}_1$ es insesgado.

Sesgo $\hat{A}_2 = b(\hat{A}_2) - A = \varepsilon\{x_1\} - A = A - A = 0.$

$\hat{A}_2$ es insesgado.

Varianza Estimadores.

$\hat{A}_1 = Var(\hat{A}_1) = Var\left(\frac{1}{N}\sum_{n=1}^{N} x_n\right)$

Dado $x_n = A + w_n$, y w_n ruido Gaussiano σ^2.

$$= \frac{1}{N^2}\sum_{n=1}^{N} Var(x_n) = \sum_{n=1}^{N} \sigma^2 = \frac{N\sigma^2}{N^2\cancel{\sigma^2}} = \frac{\sigma^2}{N}$$

Varianza $\hat{A}_2 = Var(\hat{A}_2) = Var(x_1) = \sigma^2.$

$Var(\hat{A}_1) = \frac{\sigma^2}{N}$ y $Var(\hat{A}_2) = \sigma^2$

$Var(\hat{A}_1) < Var(\hat{A}_2)$, $\frac{\sigma^2}{N} < \sigma^2$ par $N > 1$. $<$ MSE

6. Demuestre que:

$\mathcal{E}\{\mu_M\} = \mu$ con $x_n \sim N(x_n | \mu, \sigma^2)$

$$= \mathcal{E}\left(\frac{1}{N}\sum_{n=1}^{N} x_n\right) \qquad \mathcal{E}\{x_n\} = \mu$$

$$= \frac{1}{N}\sum_{n=1}^{N} \mathcal{E}\{x_n\}$$

$$= \mu$$

$\mathcal{E}\{\sigma^2_{ML}\} = \left(\frac{N-1}{N}\right)\sigma^2$ ¿Cuál debería ser la corrección sobre σ^2_{ML} para evitar el sesgo del estimador?

$$= \mathcal{E}\left(\frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{\mu}_{ML})^2\right)$$

$$= \frac{1}{N}\sum_{n=1}^{N} \mathcal{E}\{x_n^2 - 2x_n\hat{\mu}_{ML} + \hat{\mu}_{ML}\}$$

$$= \frac{1}{N}\sum_{n=1}^{N} (\sigma^2 + \mu^2 - \mathcal{E}\{\mu^2_{ML}\} \qquad Var(\mu_M) = \frac{\sigma^2}{N}$$

$$= \frac{N-1}{N}\sigma^2$$

Dado que $\mathcal{E}\{\hat{\sigma}^2_{ML}\} < \sigma^2$, el estimador σ^2_{ML} está sesgado hacia abajo. La corrección es multiplicar por $\frac{N}{N-1}$

$$\mathcal{E}\{\sigma^2_{corr}\} = \sigma^2.$$

$$\sigma^2_{corr} = \frac{N}{N-1}\sigma^2_{ML}$$

7. Demuestre $S^2 = \Lambda$

Descomposición Valores singulares (SVD):

$$\Phi = USV^T$$

Donde:

Φ es matriz descompone, U y V matrices unitarias, S matriz diagonal con valores singulares,

$V^{-1} = V^*$

$\Phi^T\Phi = (USV^*)^T (USV^*)$

$\Phi^T\Phi = VS^TU^T USV$ $\qquad U^TU = I.$

$\Phi^T\Phi = VS^TSV^*$

$\Phi^T\Phi = V\Delta V^*$ $\qquad$ Se concluye $S^TS = \Delta$

En términos valores diagonal $S^2 = \Delta$

8. Repet Min cuadrados regularizados.

$e_{cmr}(t, \Phi w) = \mathbb{E}\{\|t - \Phi w\|_2^2\} + \lambda\|w\|_2^2$

$\hat{w} = \arg\min_{w} \|y - \Phi w\|_2^2 + \lambda \|w\|_2^2$

con $\lambda \in \mathbb{R}^+$

$\langle y - \Phi w, y - \Phi w\rangle + \lambda\langle w, w\rangle$

$= y^Ty - y^T\Phi w - (\Phi w)^T y + (\Phi w)^T \Phi w + \lambda w^T w.$

$= y^Ty - 2y^T\Phi w + w^T\Phi^T\Phi w + \lambda w^T w$

Derivando

$\frac{\partial}{\partial w}(y^Ty - 2y^T\Phi w + w^T\Phi^T\Phi w + \lambda w^T w)$

$(2\Phi^T\Phi + 2\lambda I) w = 2\Phi^T y.$

$$\hat{w} = (\Phi^T\Phi + \lambda I)^{-1}\Phi^T y.$$

9. W maximizan el log-MAP.

$$p(w|y,\phi,\sigma_n^2,\sigma_w^2) \propto p(y|\phi,w,\sigma_n^2)\, p(w|\sigma_w^2)$$

donde $p(y|\phi,w,\sigma_n^2) = \prod N(y_n|0,\sigma_n^2)$

y $p(w|\sigma_w^2) = \prod N(w_n|0,\sigma_w^2)$

w que maximiza log-MAP

$$\log(p(y|\phi,w,\sigma_n^2)\, p(w|\sigma_w^2))$$

$$= \log\left(\prod_{n=1}^{N} N(y_n|0,\sigma_n^2) \prod_{n=1}^{N} N(w_n|0,\sigma_w^2)\right)$$

~~Donde~~ Log suma por propiedad.

$$= -\frac{N}{2}\left(\log(2\pi) + \log(\sigma_n^2)\right) - \frac{1}{2\sigma_n^2}\|y-\phi w\|_2^2$$

$$- \frac{Q}{2}\left[\log(2\pi) + \log(\sigma_w^2)\right] - \frac{\|w\|}{2\sigma_w^2}$$

$$= \frac{1}{2\sigma_n^2}\|y-\phi w\|_2^2 - \frac{\|w\|}{2\sigma_w^2} + cte.$$

Ahora queremos maximar.

$$\frac{\partial}{\partial w}\left(-\frac{1}{2\sigma_n^2}\|y-\phi w\|_2^2 - \frac{\|w\|}{2\sigma_w^2} + cte\right) = 0.$$

$$-\phi^T y + \left(\phi^T\phi + \frac{\sigma_n^2}{\sigma_w^2} I\right) w = 0$$

$$\hat{w}_{MAP} = \left(\phi^T\phi + \frac{\sigma_n^2}{\sigma_w^2} I\right)\phi^T y.$$

- Relación Mínimos cuadrados con MAP

Con MAP caso particular Mínimos cuadrados

$$\lambda = \frac{\sigma_n^2}{\sigma_w^2}$$

¿Qué sucede si se cambia prior?

La respuesta cambia, es necesario asumir que prior sigue distribución Gaussiana

- Descomposición **SVD**

$$\hat{w}_{MAP} = \left(V \Delta V^* + \frac{\sigma_n^2}{\sigma_w^2} I\right)^{-1} (USV^*)^T y$$

$$= V\left(S^2 + \frac{\sigma_n^2}{\sigma_w^2} I\right)^{-1} U^* V S^* V y$$

$$= V\left(S^2 + \frac{\sigma_n^2}{\sigma_w^2} I\right)^{-1} S^* V y.$$