Machine learning 236756

Assignment 1

Ron Sailer

ID: 313341596

Due date: 4.4.2017

1. P(forged)=0.001, P(fair)=0.999=P($forged^C$)
   Coin tosses of the same coin are independent of one another.

   The probability of X heads given Y tosses is binomially distributed with parameters $n = Y, p = P(heads)$. Therefore:
   P(X heads | Y tosses) = $\binom{Y}{X} \cdot P(heads)^X \cdot \left(1 - P(heads)\right)^{Y-X}$

   Let's mark the events as following:
   - A = the coin is forged.
   - B = given 10 tosses, 9 are heads.

   We are looking for $P(A|B)$.

$$P(A|B) \underbrace{=}_{Bayes} \frac{P(B|A) \cdot P(A)}{P(B)} \underbrace{=}_{Law\ of\ total\ probability} \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^C) \cdot P(A^C)}$$

$$= \frac{\binom{10}{9} \cdot 0.9^9 \cdot 0.1 \cdot \frac{1}{1000}}{\binom{10}{9} \cdot 0.9^9 \cdot 0.1 \cdot \frac{1}{1000} + \binom{10}{9} \cdot 0.5^9 \cdot 0.5 \cdot \frac{999}{1000}} = 0.03819$$

2. The number of girls is geometrically distributed with $p = 0.5$ (assuming the probability to give birth to a boy equals the probability to give birth to a girl).

   We'll mark Y as the number of girls.
   The expectancy of Y is the expected number of "failures" (girls) until "success" (boy) in an experiment with probability of success being 0.5.

   Therefore:

$$E(Y) = \frac{1-p}{p} = \frac{1-0.5}{0.5} = \frac{0.5}{0.5} = 1$$

   Because in every family there will be exactly one boy and thus $E(boys) = 1$ (The probability of there being only girls is infinitesimally small), in every family the expectancy of boys is 1. From the math, we've concluded that the expectancy of girls is also 1. Therefore, we can infer from this result that if there are enough people in the village, the difference between the number of girls and boys is most likely to be negligible. In other words, the number of boys and girls is (approximately) EQUAL.

3. a. Binomial:

   Our function which we want to maximize is $f(x_i) = \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}$.

$$L(\Theta) = \prod_{i=1}^{n} p(x_i|p) = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}$$
$$= \prod_{i=1}^{n} \left(\frac{n!}{x_i!\,(n-x_i)!} p^{x_i}(1-p)^{n-x_i}\right)$$

   Because logarithmic functions are monotonous, extremums of the Likelihood function are on the same points as the extremums of the log of the Likelihood function.

$$\ln L(\Theta) = \ln\left(\prod_{i=1}^{n} \left(\frac{n!}{x_i!\,(n-x_i)!} p^{x_i}(1-p)^{n-x_i}\right)\right)$$
$$= \sum_{i=1}^{n} \ln\left(\frac{n!}{x_i!\,(n-x_i)!} p^{x_i}(1-p)^{n-x_i}\right)$$
$$= \sum_{i=1}^{n} \left(\ln\left(\frac{n!}{x_i!\,(n-x_i)!}\right) + x_i \ln(p) + (n-x_i)\ln(1-p)\right)$$

   In order to find the extremum, we need to derive the function and compare it to zero.

$$\frac{\partial l}{\partial p} = \sum_{i=1}^{n} \left(\frac{x_i}{p} - \frac{(n-x_i)}{1-\mathrm{p}}\right) = 0$$
$$\Rightarrow \sum_{i=1}^{n} \left(\frac{x_i(1-p) - (n-x_i)p}{p(1-p)}\right) = 0$$
$$\Rightarrow \frac{\sum_{i=1}^{n}(x_i(1-p) - (n-x_i)p)}{p(p-1)} = 0$$
$$\Rightarrow \frac{\sum_{i=1}^{n}(x_i - np)}{p(p-1)} = 0$$
$$\Rightarrow \sum_{i=1}^{n}(x_i - np) = 0$$
$$\Rightarrow -n^2 p \Sigma_{i=1}^{n} x_i = 0$$
$$\Rightarrow p = \frac{\Sigma_{i=1}^{n} x_i}{n^2}$$

b. Normal / Gaussian distribution:

$$f(x_i) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

The Likelihood function is:

$$L(\Theta) = \prod_{i=1}^{n} p(x_i|\mu, \sigma^2)$$

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

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

Therefore, the log-Likelihood function is:

$$l(\Theta) = lnL(\Theta) = \ln\left((2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\Sigma_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}\right)$$

$$= \ln\left((2\pi\sigma^2)^{-\frac{n}{2}}\right) + \ln\left(e^{-\frac{\Sigma_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}\right)$$

$$= -\frac{n}{2}(\ln(2) + \ln(\pi) + \ln(\sigma^2)) - \frac{\Sigma_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

We want to find the maximum likelihood estimation, in order to do that, we need to derive the log-Likelihood function and find the extremum. (since the log function is monotonically increasing, the maximum of the Likelihood function will be on the same point as that of the log-Likelihood function.)

$$\frac{\partial l}{\partial \mu} = \frac{\Sigma_{i=1}^{n}(x_i-\mu)}{\sigma^2} = 0 \Rightarrow \Sigma_{i=1}^{n}(x_i-\mu) = 0 \Rightarrow \mu = \frac{\Sigma_{i=1}^{n}x_i}{n}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\Sigma_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} = 0 \Rightarrow \frac{n}{2\sigma^2} = \frac{\Sigma_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}$$

$$\Rightarrow n\sigma^2 = \Sigma_{i=1}^{n}(x_i-\mu)^2 \Rightarrow \sigma^2 = \frac{\Sigma_{i=1}^{n}(x_i-\mu)^2}{n}$$

c. Our function is: $f(x_i) = \frac{\lambda^{x_i}}{x_i!}e^{-\lambda}$

Likelihood:

$$L(\Theta) = \prod_{i=1}^{n} p(x_i|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}e^{-\lambda} = \frac{\lambda^{\Sigma_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}e^{-n\lambda}$$

$$\ln L(\Theta) = \ln\left(\frac{\lambda^{\Sigma_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}e^{-n\lambda}\right) = \ln\left(\lambda^{\Sigma_{i=1}^{n}x_i}\right) + \ln\left(e^{-n\lambda}\right) - \ln\left(\prod_{i=1}^{n}x_i!\right)$$

$$= \Sigma_{i=1}^{n}x_i \cdot \ln(\lambda) - n\lambda - \Sigma_{i=1}^{n}\ln(x_i!)$$

We want to find the extremum. In order to do that, let's derive the function and find the λ for which $\frac{\partial \ln L}{\partial \lambda} = 0$

$$\frac{\partial \ln L}{\partial \lambda} = \frac{\Sigma_{i=1}^{n}x_i}{\lambda} - n = 0 \Rightarrow \lambda = \frac{\Sigma_{i=1}^{n}x_i}{n}$$

4. Given that $x \sim N_2(\mu, \Sigma), x = \binom{x_1}{x_2}, \mu = \binom{\mu_1}{\mu_2}, \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

   Let's define $Z \sim N(0,1), z = \binom{z_1}{z_2}, \mu = \binom{0}{0}, z_1 = \frac{x_1-\mu_1}{\sigma_1}, z_2 = \frac{x_2-\mu_2}{\sigma_2}$

   The marginal distribution is, without loss of generality, defined as:

$$f_{Z_1}(z_1) = \int_{-\infty}^{\infty} f(z_1, z_2) dz_2 \, \sigma_2$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right) dz_2 \cdot \sigma_2$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right) dz_2$$

   Using the hint, and applying $a = \frac{1}{2(1-\rho^2)}, b = -\frac{\rho z_1}{(1-\rho^2)}, c = \frac{z_1^2}{2(1-\rho^2)}$ we get:

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \cdot \sqrt{2\pi(1-\rho^2)} \cdot \exp\left(\frac{\rho^2 z_1^2}{2(1-\rho^2)} - \frac{z_1^2}{2(1-\rho^2)}\right)$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \cdot \sqrt{2\pi(1-\rho^2)} \cdot \exp\left(-\frac{z_1^2}{2}\right)$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \cdot \sqrt{2\pi(1-\rho^2)} \cdot \exp\left(-\frac{x_1 - \mu_1}{2\sigma_1}\right)$$

   This is the formula for a variable with a distribution of $N(\mu_1, \sigma_1)$ and therefore, $x_1 \sim N(\mu_1, \sigma_1)$.

   The same calculation can be applied to show that $x_2 \sim N(\mu_2, \sigma_2)$.

b) Recall that $f(X_1|X_2 = x_2) = \frac{f(x_1,x_2)}{f_{X_2}(x_2)}$.

Since $X_2 \sim N(\mu_2, \sigma_2^2)$:

$$f(X_1|X_2 = x_2) = \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right)}{\frac{1}{\sigma_2\sqrt{2\pi}}\exp\left(-\frac{z_2^2}{2}\right)}$$

$$= \frac{\frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)\right)}{\exp\left(-\frac{z_2^2}{2}\right)}$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}}\exp\left(-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2) + \frac{z_2^2}{2}\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}}\exp\left(-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + \rho^2 z_2^2)\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}}\exp\left(-\frac{1}{2(1-\rho^2)} \cdot (z_1 - \rho z_2)^2\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}}\exp\left(\frac{-\left(\frac{x_1-\mu_1}{\sigma_1} - \rho\,\frac{x_2-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}}\exp\left(\frac{-\left(x_1 - \mu_1 - \frac{\rho\sigma_1(x_2-\mu_2)}{\sigma_2}\right)^2}{2\sigma_1^2(1-\rho^2)}\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}}\exp\left(\frac{-\left(x_1 - \left(\mu_1 + \frac{\rho\sigma_1(x_2-\mu_2)}{\sigma_2}\right)\right)^2}{2\sigma_1^2(1-\rho^2)}\right)$$

Therefore, $X_1|X_2 \sim N\left(\mu_1 + \frac{\rho\sigma_1(x_2-\mu_2)}{\sigma_2}, \sigma_1^2(1-\rho^2)\right)$

5. We can define $E(XY)$ as an inner product $\langle X,Y\rangle$.
We need to prove this is indeed an inner product, i.e. belonging to the inner product space. Therefore, it needs to satisfy the following three:

- symmetry: clearly, $\langle X,Y\rangle = E(XY) = E(YX) = \langle Y,X\rangle$
- linearity in the first argument: $\langle \alpha X,Y\rangle = E(\alpha XY) = \alpha E(XY) = \alpha\langle X,Y\rangle$
- positive-definiteness:

$$\langle X,X\rangle = E(XX) = E[X^2] \geq 0$$
$$\langle X,X\rangle = E(XX) = 0 \Leftrightarrow X = 0$$

Cauchy-Schwarz's inequality theorem states that for all vectors u,v of an inner product space, it is true that:

$$|\langle u,v\rangle|^2 \leq \langle u,u\rangle \cdot \langle v,v\rangle$$

Therefore:

$$|Cov(X,Y)|^2 = \left|E\left((X-\mu_X)\left(Y-\mu_y\right)\right)\right|^2$$
$$= \left|\left\langle X-\mu_x, Y-\mu_y\right\rangle\right|^2 \underbrace{\leq}_{C.S} \langle X-\mu_x, X-\mu_x\rangle\left\langle Y-\mu_y, Y-\mu_y\right\rangle$$
$$= E[(X-\mu_x)^2]E\left[\left(Y-\mu_y\right)^2\right] = Var(X)Var(Y)$$

Hence:

$$\left(\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}\right)^2 \leq 1$$
$$\Downarrow$$
$$-1 \leq \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \leq 1$$
$$\Downarrow$$
$$-1 \leq \rho \leq 1$$