Machine Learning Assignment # 3
Universität Bern

Due date: 10/10/2018

Late submissions will incur a penalty. Submit your answers in ILIAS (as a pdf or as a picture of your written notes if the handwriting is very clear). Submission instructions will be provided via email. You are not allowed to work with others.

Probability theory review **[Total 100 points]**

Solve each of the following problems and show all the steps of your working.

1. Show that $var[X] = E[X^2] - E[X]^2$. **[10 points]**
   Solution

$$\begin{aligned}
var[X] &= E[(X - E[X])^2] \\
&= E[X^2 - 2XE[X] + E[X]^2] \\
&= E[X^2] - E[2XE[X]] + E[E[X]^2] \\
&= E[X^2] - 2E[XE[X]] + E[X]^2 \\
&= E[X^2] - 2E[X]E[X] + E[X]^2 \\
&= E[X^2] - 2E[X]^2 + E[X]^2 \\
&= E[X^2] - E[X]^2
\end{aligned}$$

2. Show that the variance of a sum is $var[X+Y] = var[X] + var[Y] + 2cov[X,Y]$, where $cov[X,Y]$ is the covariance between X and Y. **[10 points]**
   Solution

$$\begin{aligned}
var[X+Y] &= E[((X+Y) - E[X+Y])^2] \\
&= E[((X+Y - E[X] - E[Y]))^2] \\
&= E[((X - E[X]) + (Y - E[Y]))^2] \\
&= E[(X - E[X])^2 + 2(X - E[X])(Y - E[Y]) + (Y - E[Y])^2] \\
&= E[(X - E[X])^2] + E[2(X - E[X])(Y - E[Y])] + E[(Y - E[Y])^2] \\
&= E[(X - E[X])^2] + 2E[(X - E[X])(Y - E[Y])] + E[(Y - E[Y])^2] \\
&= var[X] + 2cov[X,Y] + var[Y]
\end{aligned}$$

3. Show that the covariance matrix is always symmetric and positive semidefinite. **[10 points]**
   Solution
   The $(i,j)^{th}$ element of the covariance matrix Σ is given by

$$\Sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = E[(X_j - \mu_j)(X_i - \mu_i)] = \Sigma_{ji}$$

   so that the covariance matrix is symmetric.

For an arbitrary vector u,

$$
\begin{aligned}
u^T\Sigma u &= u^T E[(X-\mu)(X-\mu)^T]u = E[u^T(X-\mu)(X-\mu)^T u] \\
&= E[((X-\mu)^T u)^T (X-\mu)^T u] = E[((X-\mu)^T u)^2] \geq 0
\end{aligned}
$$

so that the covariance matrix is positive semidefinite.

4. Show that the uniform distribution $f(x)$ integrates to 1 **[10 points]**

$$
f(x) = \begin{cases} \frac{1}{b-a}, & \text{for } a \leq x \leq b \\ 0, & \text{otherwise.} \end{cases}
$$

Solution

$$
\begin{aligned}
\int_{-\infty}^{+\infty} f(x) = \int_a^b f(x) &= \int_a^b \frac{1}{b-a} dx \\
&= \frac{1}{b-a} \int_a^b dx = \frac{1}{b-a} \Big|_a^b x \\
&= \frac{1}{b-a}(a-b) = 1
\end{aligned}
$$

5. Show that the exponential distribution $f(x)$ integrates to 1 **[15 points]**

$$
f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } 0 \leq x < +\infty \\ 0, & \text{otherwise.} \end{cases}
$$

Solution

$$
\begin{aligned}
\int_{-\infty}^{+\infty} f(x) &= \int_0^{+\infty} f(x) dx \\
&= \int_0^{+\infty} \lambda e^{-\lambda x} dx \\
&= \lambda \int_0^{+\infty} e^{-\lambda x} dx
\end{aligned}
$$

Lets define $y = -\lambda x$ then $dy = -\lambda dx$ and $dx = -\frac{dy}{\lambda}$ (note that integration interval will change for y)

$$
\begin{aligned}
\int_{-\infty}^{+\infty} f(x) &= \lambda \int_{-\infty}^{0} e^y (-\frac{dy}{\lambda}) \\
&= -\int_{-\infty}^{0} e^y dy \\
&= -\Big|_{-\infty}^{0} e^y \\
&= -(\lim_{y \to -\infty} e^y - e^0) \\
&= e^0 - \lim_{y \to -\infty} e^y \\
&= 1 - 0 = 1
\end{aligned}
$$

6. Let $X_1, X_2, ..., X_n$ be i.i.d. Poisson random variables, with $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$.
Find the λ that maximizes the likelihood of $X_1, ..., X_n$. **[15 points]**
Solution
$l(\lambda) = \sum_{i=1}^n (x_i \log\lambda - \lambda - \log x_i!) = \log\lambda \sum_{i=1}^n x_i - n\lambda - \sum_{i=1}^n \log x_i!$
$l'(\lambda) = \frac{1}{\lambda}\sum_{i=1}^n x_i - n = 0 \rightarrow \lambda = \sum_{i=1}^n x_i / n$

7. $X \in R^n$ and $Y \in R^m$ are independent random variables. Their expectations and covariances are $E[X] = 0$, $\text{Cov}[X] = I$, $E[Y] = \mu$ and $\text{Cov}[Y] = \sigma I$, where I is the identity matrix of the appropriate size and σ is scalar. What are the expectation and covariance of the random variable $Z = AX + Y$, where $A \in R^{m \times n}$? **[15 points]**

   Solution.
   The expectation of Z can be obtained from the definition by applying the linearity of expectation,

$$E[Z] = E[AX + Y] = AE[X] + E[Y] = 0 + \mu = \mu. \tag{1}$$

   The covariance of Z is $Cov[Z] = E[ZZ^T] - E[Z]E[Z] = E[ZZ^T] - \mu\mu^T$. Substituting the definition of Z, we get the expression below.

$$\begin{aligned} E[ZZ^T] &= E[(AX+Y)(AX+Y)^T] = && (2)\\ &= E[AXX^TA^T + YX^TA^T + AXY^T + YY^T] = && (3)\\ &= AE[XX^T]A^T + E[YX^T]A^T + && (4)\\ &\quad AE[XY^T] + E[YY^T]. && (5) \end{aligned}$$

   Here we can substitute $E[XX^T] = I$ and $E[YY^T] = \sigma I + \mu\mu^T$. Because X and Y are independent, $E[XY^T] = E[X]E[Y]^T = 0$, similarly $E[YX^T] = 0$. We get $E[ZZ^T] = AA^T + \sigma I + \mu\mu^T$, therefore $Cov[Z] = AA^T + \sigma I$.

8. Suppose X, Y are two points sampled independently and uniformly on the interval [0, 1]. **[15 points]**
   What is the expectation of the left most point between X and Y?
   For the leftmost point between X and Y use the definition of the minimum between two variables:

$$\min(x, y) = \frac{x + y - |x - y|}{2}.$$

   Solution.
   Lets start by computing the expectation of the minimum, which corresponds to the leftmost variable:

$$E[min(x,y)] = \int_0^1 \int_0^1 min(x,y)p(x,y)dxdy$$

   Since X and Y are independent, we can easily decompose the joint distribution $p(x, y) = p(x)p(y)$

$$\begin{aligned} E[min(x,y)] &= \textstyle\int_0^1 \int_0^1 min(x,y)p(x)p(y)dxdy && (6)\\ &= \textstyle\int_0^1 \int_0^1 min(x,y)dxdy && (7)\\ &= \textstyle\int_0^1 \int_0^1 \frac{x+y-|x-y|}{2} dxdy && (8)\\ &= \textstyle\frac{1}{2}\int_0^1 \int_0^1 (x+y-|x-y|)dxdy && (9) \end{aligned}$$

   For the final step, we have to decompose the double integral (9) in order to strip of the absolute value of the integrand.

$$\begin{aligned} E[min(x,y)] &= \frac{1}{2}\int_0^1 \int_0^1 (x+y-|x-y|)dxdy \\ &= \frac{1}{2}\int_0^1 \left(\int_0^y x+y-(-x+y)dx + \int_y^1 x+y-(x-y)dx \right) dy \\ &= \frac{1}{2}\int_0^1 \left(\int_0^y 2xdx + \int_y^1 2ydx \right) dy = \int_0^1 \left(\int_0^y xdx + \int_y^1 ydx \right) dy \\ &= \int_0^1 \left(\int_0^y xdx + y\int_y^1 dx \right) dy = \int_0^1 \left(\Big|_0^y \frac{x^2}{2} + y\Big|_y^1 x \right) dy \\ &= \int_0^1 \left(\frac{y^2}{2} - 0 + y(1-y) \right) dy = \int_0^1 \left(y - \frac{y^2}{2} \right) dy \\ &= \int_0^1 ydy - \frac{1}{2}\int_0^1 y^2dy = \Big|_0^1 \frac{y^2}{2} - \Big|_0^1 \frac{y^3}{6} \\ &= \frac{1}{2} - 0 - (\frac{1}{6} - 0) = \frac{1}{3} \end{aligned}$$