Deep Learning Homework Assignment 1

Solutions and Grading Policy by Keyi Liu

February 9, 2018

Problem 1 (10 points)

Given 3 Random Variables(RVs), and their joint probability mass $P(X, Y, Z) = P(X)P(Y)P(Z|X, Y)$.

(a) Show that RV X and Y are marginally independent. You can either show that $P(X, Y) = P(X)P(Y)$,

$$\begin{aligned} P(X, Y) &= \sum_Z P(X, Y, Z) = \sum_Z P(X)P(Y)P(Z|X, Y) \\ &= P(X)P(Y) \sum_Z P(Z|X, Y) = P(X)P(Y) \end{aligned} \tag{1}$$

as desired. Or, you can also show that $P(X|Y) = P(X)$, of course $P(Y|X) = P(Y)$ is equivalent,

$$\begin{aligned} P(X|Y) &= \frac{P(X, Y)}{P(Y)} = \frac{\sum_Z P(X, Y, Z)}{P(Y)} = \frac{\sum_Z P(X)P(Y)P(Z|X, Y)}{P(Y)} \\ &= \frac{P(X)P(Y)\sum_Z P(Z|X, Y)}{P(Y)} = P(X) \end{aligned} \tag{2}$$

as desired.

Gradings 5 points subtotal, correctly applying the sum rule **(3 points)**, properly simplified and get the final equation another **(2 points)**.

(b) Show that X and Y are conditionally dependent given Z. By definition, you can show that $P(X, Y|Z) \neq P(X|Z)P(Y|Z)$,

$$\begin{aligned} P(X, Y|Z) &= \frac{P(X, Y, Z)}{P(Z)} = \frac{P(Z)P(Y|Z)P(X|Y, Z)}{P(Z)} \\ &= P(X|Y, Z)P(Y|Z) \end{aligned} \tag{3}$$

thus $P(X, Y|Z) \neq P(X|Z)P(Y|Z)$ as desired.

Gradings 5 points subtotal, correctly apply the chain rule **(3 points)**, obtain the final conclusion correctly **(2 points)**. Note that the way of applying the chain rule is not unique, you will get points for other ways of applying it.

Problem 2 (10 points)

The given event (rolling a die) follows a multinomial distribution. Denote the RV which is the number of landing on different values $i(i = 1, \cdots, 6)$ as X_i, and θ_i the associated probability of each value i, then the Probability Mass Function (PMF) of this multinomial distribution is given as,

$$
\begin{aligned}
f(x_1, \cdots, x_6; n, \theta_1, \cdots, \theta_6) &= Pr(X_1 = x_1, X_2 = x_2, \cdots, X_6 = x_6) \\
&= \frac{n!}{x_1! \cdots x_6!} \theta_1^{x_1} \cdots \theta_6^{x_6}
\end{aligned} \tag{4}
$$

where $\sum_{i=1}^{6} x_i = n$.

Here, we have $n = 100$, plug in the $x_i's$ and $\theta_i's$ and we will get the probability,

$$
\begin{aligned}
&P(X_1 = 20, X_2 = 15, X_3 = 10, X_4 = 5, X_5 = 20, X_6 = 30) \\
=&\frac{100!}{20! \times 15! \times 10! \times 5! \times 20! \times 30!} \times 0.25^{20} \times 0.3^{15} \times 0.1^{10} \times 0.1^{5} \times 0.15^{20} \times 0.1^{30} = 4.53 \times 10^{-14}
\end{aligned} \tag{5}
$$

Gradings Identifying that this is a multinomial distribution **(3 points)**, correctly show the probability mass function you will get another **(5 points)**, and calculate the final probability correctly **(2 points)**.

Problem 3 (10 points)

Given the joint probability distribution of $P(X, Y, Z)$, you are asked to show that any pair of RVs are marginally independent of each other. Take RV X and Y as an example, by definition, you will need to show that $P(X, Y) = P(X)P(Y)$ for each configuration of (X, Y)(It is also correct if you can show that $P(X|Y) = P(X)$).

You can start from the marginal,

$$
\begin{aligned}
P(X = 0) &= \sum_{Y,Z} P(X = 0, Y, Z) = 1/12 + 1/6 + 1/6 + 1/12 = 1/2 \\
P(X = 1) &= 1 - P(X = 0) = 1/2
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
P(Y = 0) &= \sum_{X,Z} P(X, Y = 0, Z) = 1/12 + 1/6 + 1/6 + 1/12 = 1/2 \\
P(Y = 1) &= 1 - P(Y = 0) = 1/2
\end{aligned} \tag{7}
$$

Then, compute the joint probability for X and Y,

$$
\begin{aligned}
P(X = 1, Y = 1) &= \sum_{Z} P(X = 1, Y = 1, Z) = 1/12 + 1/6 = 1/4 \\
P(X = 0, Y = 0) &= \sum_{Z} P(X = 0, Y = 0, Z) = 1/12 + 1/6 = 1/4 \\
P(X = 1, Y = 0) &= \sum_{Z} P(X = 1, Y = 0, Z) = 1/12 + 1/6 = 1/4 \\
P(X = 0, Y = 1) &= \sum_{Z} P(X = 0, Y = 1, Z) = 1/12 + 1/6 = 1/4
\end{aligned} \tag{8}
$$

We can then check that for each configuration of (X, Y), we have $P(X, Y) = P(X)P(Y)$, thus it follows that X and Y are marginally independent of each other. You can repeat similar process for other two pair of RVs, and get the same conclusion.

Gradings You will get **9 points** if you correctly show one of the three pairs of RVs' independency, i.e. **(3 points)** for each of $P(X)$, $P(Y)$, and $P(X,Y)$ in my sample solution. If you get all three pairs correctly, you will get the total 10 points.

Problem 4 (10 points)

The covariance matrix $\mathbf{\Sigma}$ for RV $\mathbf{X}$ is a symmetric and positive semi-definite matrix with all its entries real values. Thus the Eigen Decomposition exists, and is given as,

$$\mathbf{\Sigma} = \mathbf{Q\Lambda Q^{-1}} \tag{9}$$

where $\mathbf{Q}$ is a $N \times N$ orthonormal matrix, with its columns the eigenvectors of $\mathbf{\Sigma}$, and Λ a diagonal matrix with eigenvalues of $\mathbf{\Sigma}$ on its main diagonal.

The Singular Value Decomposition is given as,

$$\mathbf{\Sigma} = \mathbf{UDV^T} \tag{10}$$

and for symmetric matrix, it is reduced to

$$\mathbf{\Sigma} = \mathbf{UDU^T} \tag{11}$$

where $\mathbf{U}$ is also a orthonormal matrix containing singular vectors in its columns, and $\mathbf{D}$ a diagonal matrix with the singular value of $\mathbf{\Sigma}$ along it main diagonal.

From Eigen Decomposition we have,

$$\begin{aligned} \mathbf{\Sigma^T\Sigma} &= (\mathbf{Q\Lambda Q^{-1}})^\mathbf{T}\mathbf{Q\Lambda Q^{-1}} \\ &= \mathbf{Q^{-T}\Lambda^T Q^T Q\Lambda Q^{-1}} \\ &= \mathbf{Q\Lambda^T\Lambda Q^T} \end{aligned} \tag{12}$$

Note that, $\mathbf{Q}$ is orthonormal, so we have $\mathbf{Q^T} = \mathbf{Q^{-1}}$.
From Singular Value Decomposition we have,

$$\begin{aligned} \mathbf{\Sigma^T\Sigma} &= (\mathbf{UDV^T})^\mathbf{T}\mathbf{UDV^T} \\ &= \mathbf{UD^TDU^T} \end{aligned} \tag{13}$$

It is obvious that both $\mathbf{\Lambda}$, and $\mathbf{D}$ contain the square root of the eigenvalues of matrix $\mathbf{\Sigma^T\Sigma}$, thus $\mathbf{\Lambda} = \mathbf{D}$, and it follows that $\mathbf{U} = \mathbf{Q}$, namely, the Eigen Decomposition is equivalent to SVD.

Gradings Correctly give the formula of both Eigen Decomposition and SVD worth **(4 points)**, compare the resulting matrices and give proper analysis **(6 points)**, specifically, claim that in $\mathbf{UDV^T}$, $\mathbf{U} = \mathbf{V}$, $\mathbf{\Lambda}$ and $\mathbf{D}$ are diagonal matrices containing eigenvalues, and that $\mathbf{Q} = \mathbf{U}$, each **(2 points)**.

Problem 5 (10 points)

Given a scalar functional $f(\mathbf{X})$, with the coefficient matrices A, b, C, d not a function of $\mathbf{X}$, let

$$g(\mathbf{X}) = (A\mathbf{X} + b)^T(C\mathbf{X} + d) + \lambda\mathbf{X}^T\mathbf{X} \tag{14}$$

Thus we shall calculate the partial derivative w.r.t. $\mathbf{X}$, i.e.

$$\begin{aligned}\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} &= \frac{\partial \log(g(\mathbf{X}))}{\partial g(\mathbf{X})} \times \frac{\partial g(\mathbf{X})}{\partial \mathbf{X}} \\ &= \frac{1}{g(\mathbf{X})} \times \frac{\partial g(\mathbf{X})}{\partial \mathbf{X}}\end{aligned} \tag{15}$$

where

$$\begin{aligned}\frac{\partial g(\mathbf{X})}{\partial \mathbf{X}} &= \frac{\partial[(A\mathbf{X} + b)^T(C\mathbf{X} + d)]}{\partial \mathbf{X}} + \lambda\frac{\partial \mathbf{X}^\mathbf{T}\mathbf{X}}{\partial \mathbf{X}} \\ &= \frac{\partial(A\mathbf{X} + b)}{\partial \mathbf{X}}(C\mathbf{X} + d) + \frac{\partial(C\mathbf{X} + d)}{\partial \mathbf{X}}(A\mathbf{X} + b) + 2\lambda\frac{\partial \mathbf{X}}{\partial \mathbf{X}}\mathbf{X} \\ &= A^T(C\mathbf{X} + d) + C^T(A\mathbf{X} + b) + 2\lambda\mathbf{X}\end{aligned} \tag{16}$$

Thus combine Equation (14), (15), and (16), we have,

$$\frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \frac{A^T(C\mathbf{X} + d) + C^T(A\mathbf{X} + b) + 2\lambda\mathbf{X}}{(A\mathbf{X} + b)^T(C\mathbf{X} + d) + \lambda\mathbf{X}^T\mathbf{X}} \tag{17}$$

Note that, the dimension of the result is the same as $\mathbf{X}^{N\times 1}$.

The rules that are used here,

(1)

$$\frac{\partial U\mathbf{X}}{\partial \mathbf{X}} = U^T$$

(2)

$$\frac{\partial U^T V}{\partial \mathbf{X}} = \frac{\partial U}{\partial \mathbf{X}}V + \frac{\partial V}{\partial \mathbf{X}}U$$

Gradings Correctly applying the chain rule and get the correct derivative for $\log(\cdot)$ **(5 points)**, correctly applying the above rules and calculate the remaining parts of the derivatives **(5 points)**.