Group Assignment 3

Introduction to Data Science 1MS041

Truls Karlsson, Nathalie Borglund, Noah Wassberg,
Andreas Larsson

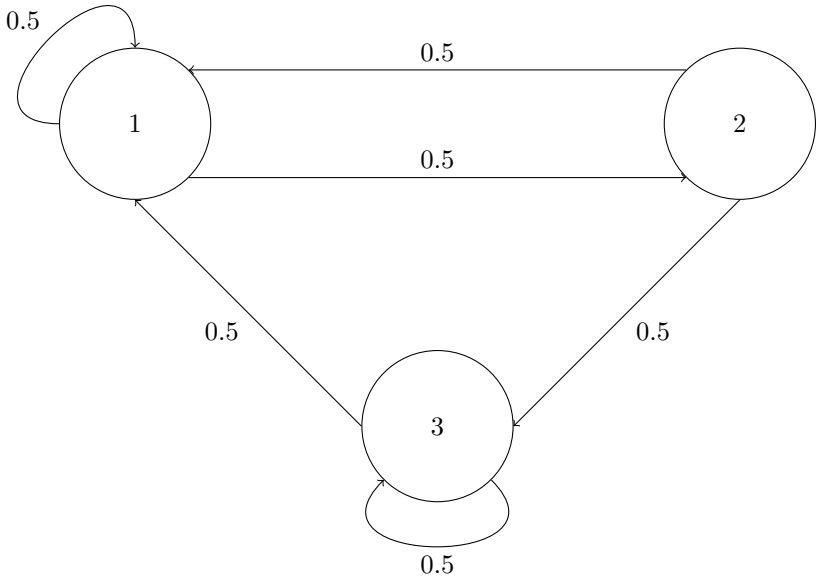
December 2024

1.

Consider a three state (1, 2, 3) Markov chain with transition matrix $P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix}$

a

Question: Draw the transition diagram.
Answer:



b

Question: Find stationary distribution π.
Answer: First we write the problem on the form $\pi = \pi P$ which can be rewritten as $\pi^T = P^T \pi^T$. Now we can interpret it as $P^T \pi^T = \lambda \pi^T$ where $\lambda = 1$. The problem is now reduced to finding the eigenvector corresponding to $\lambda = 1$.

To find the eigenvector with $\lambda = 1$ we first calculate the eigenvalues to verify that there is a $\lambda = 1$.

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix} \iff P^T = \begin{pmatrix} 0.5 & 0.5 & 0.5 \\ 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 \end{pmatrix}$$

Then we solve $\det(P^T - \lambda I) = 0$ to get the eigenvalues.

$$\left| \begin{pmatrix} 0.5 & 0.5 & 0.5 \\ 0.5 & 0 & 0 \\ 0 & 0.5 & 0.5 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right| = \left| \begin{pmatrix} 0.5 - \lambda & 0.5 & 0.5 \\ 0.5 & -\lambda & 0 \\ 0 & 0.5 & 0.5 - \lambda \end{pmatrix} \right|$$

$$\begin{aligned} &(0.5 - \lambda)(-\lambda(0.5 - \lambda) - 0 * 0.5) + \\ &= 0.5(0 * 0 - 0.5(0.5 - \lambda)) + \\ &\quad 0.5(0.5 * 0.5 - (-\lambda) * 0) \end{aligned}$$

$$\begin{aligned} &(0.5 - \lambda)(\lambda^2 - 0.5\lambda) + \\ &= 0.5(0.5\lambda - 0.25) + \\ &\quad 0.5 * 0.25 \end{aligned}$$

$$= (0.5 - \lambda)(\lambda^2 - 0.5\lambda) + 0.25\lambda$$

$$= -\lambda^3 + \lambda^2$$

$$= \lambda^2(1 - \lambda)$$

$$\implies \lambda_1 = 1, \lambda_{2,3} = 0$$

Now we have determined that there is an eigenvalue equal to one for our Markov Chain and we can now calculate the eigenvector for $\lambda = 1$. We do this by solving $(P^T - \lambda I) = \vec{v_{\lambda_1}}$ by Gaussian elimination.

$$\begin{bmatrix} -0.5 & 0.5 & 0.5 \\ 0.5 & -1 & 0 \\ 0 & 0.5 & -0.5 \end{bmatrix} \xrightarrow{R_1+1R_3} \begin{bmatrix} -0.5 & 1 & 0 \\ 0.5 & -1 & 0 \\ 0 & 0.5 & -0.5 \end{bmatrix}$$
$$\xrightarrow{R_1+1R_2} \begin{bmatrix} 0 & 0 & 0 \\ 0.5 & -1 & 0 \\ 0 & 0.5 & -0.5 \end{bmatrix}$$
$$\xrightarrow{2R_3} \begin{bmatrix} 0 & 0 & 0 \\ 0.5 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

This results in the equations:

$$\begin{cases} 0.5v_1 - v_2 = 0 \\ v_2 - v_3 = 0 \end{cases} \implies \begin{cases} v_1 = 2v_2 \\ v_2 = v_3 \end{cases}$$

We parametrize v_3 as t and get the eigenvector: $\vec{v_{\lambda_1}} = \begin{pmatrix} 2t \\ t \\ t \end{pmatrix}, \forall t \in \mathbb{R}$

Since we are looking for a stationary distribution the vector must sum to 1. To find a t that satisfies this we solve the equation: $1 = 2t + t + t \iff 1 = 4t \iff t = 0.25$

This gives the stationary distribution: $\pi = \begin{pmatrix} 0.5 \\ 0.25 \\ 0.25 \end{pmatrix}$ ■

c

Question: Given that the chain is in state 1 at time 1, what is the probability that the chain is in state 2 at time 4?

Answer: From the question we can derive that the initial state, let us denote it as s_0, can be represented as a vector with a one or zero representing each state. $s_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Given is the transition matrix. For each timestep we multiply the current state with the transition matrix to get the next step. By doing this four times we get the probabilities of the chain being in the different states at timestep 4. The calculation is: $Prob_{t=4}(((s_0P)P)P)P = s_0P^4$ and by inserting the values we get:

$$
\begin{aligned}
Prob_{t=4} &= \begin{pmatrix}1\\0\\0\end{pmatrix}\begin{pmatrix}0.5 & 0.5 & 0\\0.5 & 0 & 0.5\\0.5 & 0 & 0.5\end{pmatrix}^4 \\
&= \begin{pmatrix}1\\0\\0\end{pmatrix}\begin{pmatrix}0.5 & 0.25 & 0.25\\0.5 & 0.25 & 0.25\\0.5 & 0.25 & 0.25\end{pmatrix} \\
&= \begin{pmatrix}0.5\\0.25\\0.25\end{pmatrix}
\end{aligned}
$$

As mentioned before, the resulting vector is the probabilities of being in the different states (1,2,3) at time four after starting in state one. The middle element of the vector represent state 2 and the answer to the question is therefore 0.25.

d

Question: Given that the chain is in state 1 at time 1, what is the expected time until the chain is in state 3 the first time?
Answer: First we observe how the chain transitions from state 1 and state 2 from the transition matrix or diagram. It is evident that the chain can make two choices at each state and we will denote the sum of both probabilities as the expectation at state one (E_1) and two (E_2) respectively. At state one the chain may transition to state one (loop) or to state two. At state two the chain may transition to state one or three. When translating this into equations we get a system which we then solve for E_1:

$$
\begin{cases} E_1 &= 0.5(1+E_1)+0.5(1+E_2) \\ E_2 &= 0.5(1)+0.5(1+E_1) \end{cases} \iff \begin{cases} E_1 &= 2+E_2 \\ E_2 &= 1+0.5E_1 \end{cases}
$$

$$
\implies E_1 = 2+1+0.5E_1 \iff 2E_1 = 6+E_1
$$

$$
\iff E_1 = 6 \implies E_2 = 4
$$

The expected time to reach state three for the first time after starting in state one is six.

e

Question: What is the period of each state?
Answer: We examine the greatest common divisor (GCD) of the lengths of all possible transitions from S_i back to S_i. For S_1 there are the following:

1. $S_1 -> S_1, l = 1$
2. $S_1 -> S_2 -> S_1, l = 2$
3. $S_1 -> S_2 -> S_3 -> S_1, l = 3$
4. $S_1 -> S_2 -> S_3 -> (loop\ at\ S_3) -> S_1, l = 4, 5, 6, ...$

Since the fourth option goes on infinitely the GCD of state 1 is 1. If the GCD of one state is 1 then the GCD of all states will we 1 and the period is 1.

2

no, that is not correct. Periods of states can be different, you need to

Assume that we are trying to classify a binary outcome Y, i.e. our data is of the form $(X, Y) \sim F_{X,Y}$, where $Y \in \{0, 1\}$ and $X \in \mathbb{R}^d$. We have used data to train a classifier $g(X)$. We can evaluate the performance of the classifier using i.i.d. testing data, $(X_1, Y_1), ..., (X_n, Y_n)$. We are interested in estimating the following quantities:
Precision: $P(Y = 1|g(X) = 1)$,
Recall: $P(g(X) = 1|Y = 1)$.

a

Question: Write down the empirical version of the precision and recall.
Answer: Precision is defined as the number of instances when the classifier correctly predicts a one, divided by the number of all times the classifier predicts a one. Recall on the other hand is again the number of instances when the classifier correctly predicts a one, but now divided by the number of true labels that are equal to one. In the problem posed above the performance of the classifier is based on the test data and the empirical versions will then be:

not quite, you should use formula for conditional probability to express precision and

$$\mathbb{E}_{precision}[Y_i = 1|g(X_i) = 1, \quad X_i, Y_i \in \{(X_1, Y_1), ..., (X_n, Y_n)\}]$$
$$\mathbb{E}_{recall}[g(X_i) = 1|Y_i = 1, \quad X_i, Y_i \in \{(X_1, Y_1), ..., (X_n, Y_n)\}] \quad \blacksquare$$

b

Question: Let us now think that the variable Y denotes if a battery's health has deteriorated or not, and let X denote a bunch of constructed health indicators about the battery. If the model $g(X)$ predicts that the battery has deteriorated you need to run a test to confirm this. The cost of running the test is c when the battery is not deteriorated. On the other hand, if the battery is in fact deteriorated and the test is not run, the battery will die during use and the cost of this is d. Define a random variable representing the cost of the decision $g(X)$ and write down the formula for the expected cost in terms of the precision and recall.
Answer: The cost is dependent on the output of the classifier and the true label (state of the battery). There are three cases that needs to be considered when defining the cost. When the classifier predicts a false positive and an unnecessary test is run; when the classifier predicts a false negative and a test is needed but not performed; when the classifier predicts correctly and no consequences follow. The cost if therefore defined as:

$$Cost(g(X), Y) = \begin{cases} c, & \text{if } g(X) = 1 \text{ and } Y = 0 \\ d, & \text{if } g(X) = 0 \text{ and } Y = 1 \\ 0, & \text{otherwise} \end{cases}$$

Now we formulate the expression for the expected cost based on the definition above:

$$\begin{aligned}\mathbb{E}[Cost(g(X), Y)] &= c * \mathbb{P}(g(X) = 1, Y = 0) + d * \mathbb{P}(g(X) = 0, Y = 1) \\ &= c * \mathbb{P}(Y = 0|g(X) = 1) * \mathbb{P}(g(X) = 1) + d * \mathbb{P}(g(X) = 0|Y = 1) * \mathbb{P}(Y = 1)\end{aligned}$$

Notice that $\mathbb{P}(Y = 0|g(X) = 1)$ is the probability of a false positive. Let us view it as the complement of the probability a label being correct when classified as one, then we have: $\mathbb{P}(Y = 0|g(X) = 1) = 1 - \mathbb{P}(Y = 1|g(X) = 1) = 1 - \textit{precision}$. Look also closer to the term $\mathbb{P}(g(X) = 0|Y = 1)$, interpreted as false negative, which is the complement of the classifier predicting a one when the true label also is one, $\mathbb{P}(g(X) = 0|Y = 1) = 1 - \mathbb{P}(g(X) = 1|Y = 1) = 1 - \textit{recall}$. By substituting these interpretations into the expected cost expression we get:

$$\begin{aligned}\mathbb{E}[Cost(g(X), Y)] =& c * (1 - \mathbb{P}(Y = 1|g(X) = 1)) * \mathbb{P}(g(X) = 1) + \\ & d * (1 - \mathbb{P}(g(X) = 1|Y = 1)) * \mathbb{P}(Y = 1)\end{aligned}$$ ■

c

Question: Advanced question: can you produce a confidence interval for the expected cost? What about the precision and the recall?
Answer: There are several dependencies in the cost expression and below we will start to evaluate the confidence intervals using a bottom up approach. First we define the confidence intervals for precision by expanding its empirical definition with an indicator function (I) that is one when the inner expression is true and zero otherwise.

$$\textit{precision} = \frac{\Sigma_{i=1}^{n} I(Y_i = 1 \ \textit{and} \ g(X_i) = 1)}{\Sigma_{i=1}^{n} I(g(X_i) = 1)}$$

We do now want to determine on which interval the error of *precision* lies and to do that we start by defining the error as the standard error, where I again is an indicator defined as before.

$$SE_{precision} = \sqrt{\frac{\mathbb{V}}{n}} = \sqrt{\frac{\textit{precision}(1 - \textit{precision})}{\Sigma I(g(X) = 1)}}$$

We assume that that the sample size is large enough for the central limit theorem to apply and that the standard error is normally distributed. Therefore the confidence interval for precision is defined as, where $Z_{\frac{\alpha}{2}}$ is a Z-score value based on the α threshold:

$$[\textit{precision} - Z_{\frac{\alpha}{2}} * SE_{precision}, \textit{precision} + Z_{\frac{\alpha}{2}} * SE_{precision}]$$

In the same manner we define the confidence interval for recall:

$$\textit{recall} = \frac{\Sigma_{i=1}^{n} I(Y_i = 1 \ \textit{and} \ g(X_i) = 1)}{\Sigma_{i=1}^{n} I(Y_i = 1)}$$

$$SE_{recall} = \sqrt{\frac{\mathbb{V}}{n}} = \sqrt{\frac{\textit{recall}(1 - \textit{recall})}{\Sigma I(Y = 1)}}$$

$$[recall - Z_{\frac{\alpha}{2}} * SE_{recall}, recall + Z_{\frac{\alpha}{2}} * SE_{recall}]$$

Since the cost is dependent on multiple factors, which in turn have their own dependencies, an exact confidence interval is not possible. For example the four terms *precision*, *recall*, $\mathbb{P}(g(X) = 1)$ and $\mathbb{P}(Y = 1)$ all provide their own confidence intervals which may or may not be correlated and therefore the interval for the overall cost cannot be exactly determined.

3

Question: Let X and Y be two d-dimensional zero mean, unit variance Gaussian random vectors. Show that X and Y are nearly orthogonal by calculating their dot product. Can you for instance also bound the probability that the dot product is larger than ϵ?
Answer: We have the two random gaussian vectors $X = [x_1, x_2, ..., x_d]^T$ and $Y = [y_1, y_2, ..., y_d]^T$ where $E[x_i] = E[y_i] = 0$ and $Var[x_i] = Var[y_i] = 1$ for $i = 1, 2, 3...d$. To assess near orthogonality of X Y we calculate their dot product.

$$X \cdot Y = \sum_{i=1}^{d} x_i \cdot y_i$$

which also tells us

$$E[X \cdot Y] = \sum_{i=1}^{d} E[x_i \cdot y_i] = \sum_{i=1}^{d} E[x_i] \cdot E[y_i] = 0$$

and

$$Var[X \cdot Y] = \sum_{i=1}^{d} E[x_i^2 \cdot y_i^2] - (E[x_i \cdot y_i])^2 = \sum_{i=1}^{d} E[x_i^2] \cdot E[y_i^2] = \sum_{i=1}^{d} 1 = d$$

CLT tells us that $X \cdot Y \sim N(0, d)$ as $d \to \infty$. LLN states that the average of IID random variables converge to their average. This gives us that $\frac{\|X\|^2}{d} \xrightarrow{p} 1$ as $d \to \infty$, thus $\frac{\|X\|}{d} \xrightarrow{p} \sqrt{d}$ as $d \to \infty$. Or in other words as d becomes large $||X||$ becomes close to $\sqrt{d}$ with high probability. Note that this also applies to Y. Now, to assess near orthogonality we look at the normalized dot product, giving us.

$$\frac{X \cdot Y}{||X|| \cdot ||Y||} \approx \frac{X \cdot Y}{\sqrt{d}\sqrt{d}} = \frac{X \cdot Y}{d} \sim N(0, \frac{1}{d})$$

Looking at the distribution above it becomes clear that as d grows large the variance becomes near zero, meaning that the probability that the dot product being zero becomes high which tells us that the dot product is near orthogonal.We can now use the Gaussian tail bound on the probability that the dot product is larger than ϵ. Giving us

$$P\left(\left|\frac{X \cdot Y}{d}\right| > \epsilon\right) \leq 2exp\left(-\frac{\epsilon^2 d}{2}\right)$$

As $d \to \infty$ the exponential becomes more negative and thus tends to 0.

here, CLT can be used to appr

4

Let $u_1, ..., u_r$ be $n \times 1$ unit length vectors that are linearly independent, i.e. $\sum_{i=1}^{r} \alpha_i u_i = 0$ implies that $\alpha_i = 0$ for all i.

a

Question: Verify that the matrix $u_i u_i^T$ is a rank one matrix for all i. What is the null-space and range of $u_i u_i^T$?
Answer: We have the following:

$$u_i = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{in} \end{pmatrix}, \quad u_i^T = \begin{pmatrix} u_{i1} & u_{i2} & \cdots & u_{in} \end{pmatrix}.$$

Multiplying these gives:

$$u_i u_i^T = \begin{pmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{in} \end{pmatrix} \begin{pmatrix} u_{i1} & u_{i2} & \cdots & u_{in} \end{pmatrix} = \begin{pmatrix} u_{i1}u_{i1} & u_{i1}u_{i2} & \cdots & u_{i1}u_{in} \\ u_{i2}u_{i1} & u_{i2}u_{i2} & \cdots & u_{i2}u_{in} \\ \vdots & \vdots & \ddots & \vdots \\ u_{in}u_{i1} & u_{in}u_{i2} & \cdots & u_{in}u_{in} \end{pmatrix}.$$

u_i is a $n \times 1$ matrix and $u_i u_i^T$ is a $n \times n$ matrix. The rank of a matrix formed by a single outer product of two non-zero vectors is always one. Since u_i is a non-zero vector we can conclude that $u_i u_i^T$ has rank one.

The null-space consists of all vectors x such that

$$(u_i u_i^T)x = 0.$$

Since $(u_i u_i^T)x = u_i(u_i^T x)$ we need $u_i^T x = 0$. From this we get that the null-space of $u_i u_i^T$ is the set of all vectors orthogonal to u_i which is given by

$$\mathcal{N}(u_i u_i^T) = \{x \in \mathbb{R}^n | u_i^T x = 0\}$$

The range of $u_i u_i^T$ is the subspace spanned by u_i which is given by

$$Range(u_i u_i^T) = \{\alpha u_i : \alpha \in \mathbb{R}\}$$

b

Question: Verify that the matrix $U = \sum_{i=1}^{r} u_i u_i^T$ is a rank r matrix.
Answer: We have

$$U = \sum_{i=1}^{r} u_i u_i^T$$

Let $U_r = [u_1\ u_2\ ...\ u_r]$ be a $n \times r$ matrix formed by placing the u_i vectors as columns. We know that u_i is of unit length and that $\{u_i, ..., u_r\}$ is linearly independent. From this we can rewrite U as

$$U = U_r U_r^T$$

The rank of U_r is r since its columns are linearly independent and the rank of the product $U_r U_r^T$ is at the *minimum* of the ranks of U_r and U_r^T, which we already know is r. Since $u_1, ..., u_r$ are linearly independent, we can conclude that U maps an r-dimensional space onto itself and therefore is a rank r matrix.

c

i.

Question: If we perform SVD on U, are the vectors $u_1, ..., u_r$ the same as the right singular vectors? If not, can you give an example?
Answer: **No**, and we can show this with the following example. Let $n = 2$ and

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

u_1 and u_2 are linearly independent but not orthogonal. We then get

$$U = u_1 u_1^T + u_2 u_2^T = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

The eigenvectors (the signular vectors) of $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ will not just be u_1 and u_2. Instead, the SVD will give a orthonormal set that might look like the following v_1 and v_2:

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

These differ from u_1 and u_2 and our example is therefore complete.

ii.

Question: What if the vectors $u_1, ..., u_r$ are all orthogonal? In this case, what are the singular values of U?
Answer: In this case, we now assume that the vectors $u_1, ..., u_r$ are all orthogonal which means

that U is now precisely the orthogonal projection matrix onto the r-dimensional subspace spanned by these vectors. What this means is that

$$U = \sum_{i=1}^{r} u_i u_i^T$$

is an orthogonal projector. Since $u_1, ..., u_r$ are orthogonal, this means that $u_i u_j^T = 0$ for $i \neq j$. They are also unit vectors which means that $u_i u_i^T$ has eigenvalue 1 along u_i and 0 elsewhere. Hence, the singular values of U are

$$\begin{cases} 1 \text{ for each of the } r \text{ orthonormal vectors} \\ 0 \text{ for the remaining } n - r \text{ orthogonal directions} \end{cases}$$

Answer: The singular values are 1 with multiplcity r and 0 with multiplicity $n - r$.

5

We know that $X \sim \text{Uniform}(B_1)$, where B_1 is a unit ball in $\mathbb{R}^d$ and we define $Y = ||X||_2$. From this, we know that X lies within the unit ball, and that Y is the distance between X and the origin.

a

Question: Find the distribution function of Y.
Answer: For this task, we want to find the distribution of Y. Since we know that Y is the Euclidean norm of X. We know that Y can take values between 0 and 1 because X lies within the unit ball.

From this, we can assume that the point X lies within a smaller ball with radius y, where $0 \leq y \leq 1$, when $Y \leq y$. This is because Y measures the distance from the origin, and if $Y \leq y$, this means the point X must lie inside the ball with radius y. From this we can calculate the probability $P(Y \leq y)$ assuming that the volume of the ball with radius y is proportional to the unit ball B_1.

We therefore write as follows:

$$P(Y \leq y) = \frac{\text{Volume of the ball with radius y}}{\text{Volume of the unit ball}}$$

From Theorem 10.8, we know that the volume of a unit ball with dimensions d is:

$$|B_1| = \frac{2\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})}$$

To define the volume for the ball with radius y, we conclude that the volume is scaled with y^d because the radius decreases or increases equally in all d directions. This means that the volume of the ball with radius y can be written as:

$$\frac{2\pi^{\frac{d}{2}} y^d}{d\Gamma(\frac{d}{2})}$$

Inserting these into the equation, gives us:

$$P(Y \leq y) = \frac{\frac{2\pi^{\frac{d}{2}} y^d}{d\Gamma(\frac{d}{2})}}{\frac{2\pi^{\frac{d}{2}}}{d\Gamma(\frac{d}{2})}} = y^d$$

So, therefore is the CDF of Y:

$$\begin{cases} 0, & y < 0 \\ y^d, & 0 \leq y \leq 1 \\ 1, & y > 1 \end{cases}$$

b

Question: What is the distribution of $\ln(1/Y)$?
Answer: We first define Z as $Z = ln(\frac{1}{Y})$. We now want to find the CDF of Z. Which by definition is written as:

$$F_Z(z) = \mathbb{P}(Z \leq z) \quad \text{for} \quad z \in \mathbb{R}$$

We start by substitute Z into the equation above and get:

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(ln(\frac{1}{Y}) \leq z)$$

We then further simplify this as follows:

$$F_Z(z) = \mathbb{P}(ln(\frac{1}{Y}) \leq z)$$
$$F_Z(z) = \mathbb{P}(ln(1) - ln(Y) \leq z)$$

And since $ln(1) = 0$ we can write this as:

$$F_Z(z) = \mathbb{P}(-ln(Y) \leq z)$$

We then continue to multiply it with -1 on both sides, which gives us:

$$F_Z(z) = \mathbb{P}(ln(Y) \geq -z)$$

Now we remove the logarithm, by taking the exponential of both sides:

$$F_Z(z) = \mathbb{P}(Y \geq e^{-z})$$

The CDF of Y gives us the probability $\mathbb{P}(Y \leq y)$, and since we are looking for $\mathbb{P}(Y \geq e^{-z})$ we need to use the complement rule, as follows:

$$\mathbb{P}(Y \geq e^{-z}) = 1 - \mathbb{P}(Y < e^{-z})$$

From part a). do we know that the CDF of Y is $F_y(y) = y^d$ for $0 \leq y \leq 1$. By substituting this we get:

$$\mathbb{P}(Y \geq e^{-z}) = F_Y(e^{-z}) = (e^{-z})^d = e^{-zd}$$

We then substitute this back to the equation for $F_Z(z)$ and get:

$$F_Z(z) = 1 - e^{-zd}$$

This gives the CDF for $Z = ln(\frac{1}{Y})$:

$$F_Z(z) = \begin{cases} 0, & z < 0 \\ 1 - e^{-zd}, & z \geq 0 \end{cases} \quad \blacksquare$$

c

Question: Calculate $\mathbb{E}[\ln(1/Y)]$, first using the distribution function of Y and then using the distribution function of $\ln(1/Y)$.
Answer: To solve this we need to find the expectation of $g(Y) = ln(1/Y)$, where Y is our random variable. From definition 2.53, we can write the expectation of $g(Y)$ as follows:

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y) f_Y(y) \quad dy$$

From part a), we know that the CDF of Y is $F_Y(y) = y^d$ when $0 \leq y \leq 1$. From this we can see that the PDF is $f_Y(y) = dy^{d-1}$.

$$f_Y(y) = dy^{d-1}, \quad \text{for } 0 \leq y \leq 1.$$

We simplify $ln(\frac{1}{Y})$ as we did in part b). And get the integral:

$$\mathbb{E}[ln(\frac{1}{Y})] = \int_{-\infty}^{\infty} -ln(\frac{1}{Y}) dy^{d-1} \quad dy$$
$$\mathbb{E}[ln(\frac{1}{Y})] = \int_{0}^{1} -ln(y) dy^{d-1} \quad dy$$
$$\mathbb{E}[ln(\frac{1}{Y})] = -d \int_{0}^{1} ln(y) y^{d-1} \quad dy$$

We then use this rule for integration $\int u \quad dv = uv - \int v \quad du$, and get:

$$u = ln(y)$$
$$dv = y^{d-1} dy$$
$$du = \frac{1}{y} dy$$
$$v = \frac{y^d}{d}$$
$$\mathbb{E}[ln(\frac{1}{Y})] = -d \left(\left[\frac{\ln(y) y^d}{d} \right]_0^1 - \int_0^1 \frac{y^d}{d} \frac{1}{y} dy \right)$$

We continue by calculating each part separately:

$$\left[\frac{\ln(y) y^d}{d} \right]_0^1 = \left(\frac{\ln(1) 1^d}{d} - \frac{\ln(0) 0^d}{d} \right) = 0 - 0 = 0$$

$$\begin{aligned}
\int_0^1 \frac{y^d}{d}\frac{1}{y}dy &= \frac{1}{d}\int_0^1 y^d \frac{1}{y}dy \\
&= \frac{1}{d}\int_0^1 \frac{y^d}{y}dy \\
&= \frac{1}{d}\left[\frac{y^{d-1+1}}{d-1+1}\right]_0^1 \\
&= \frac{1}{d}\left[\frac{y^d}{d}\right]_0^1 \\
&= \frac{1}{d}\left(\frac{1^d}{d} - \frac{0^d}{d}\right) \\
&= \frac{1}{d} * \frac{1}{d} = \frac{1}{d^2}
\end{aligned}$$

By inserting both parts to the equation we get:

$$\mathbb{E}[ln(\frac{1}{Y})] = -d * (0 - \frac{1}{d^2}) = \frac{1}{d}$$

also using distribution of ln(1/y)