EE 641 Final Exam

December 14, Fall 2021

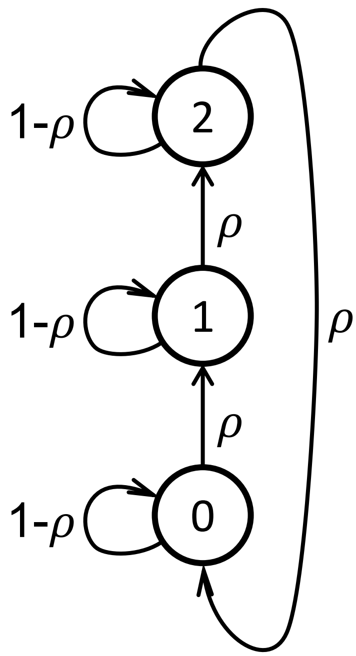
Name: **Key** ______________________________

Q1: Instructions (4pt)

Rules: I understand that this is an open book exam that shall be done within the allotted time of 180 minutes. I can use my notes, previous posted exams and exam solutions, and web resources. However, I will not communicate with any other person other than the official exam proctors during the exam, and I will not seek or accept help from any other persons other than the official proctors.

Signature: ______________________________

Q2: Markov Chain (35pt)



Let $\{X_n\}_{n=0}^{\infty}$ be a homogeneous Markov chain with states $\{0, 1, 2\}$ and state-transition diagram as shown above with $\rho \in (0, 1)$.

Q2.1:

Write out the state transition matrix, P, for X_n.

Q2.2:

Prove or disprove that X_n is irreducible.

Q2.3:

Prove or disprove that X_n is periodic.

Q2.4:

Prove or disprove that X_n is ergodic.

Q2.5:

Prove or disprove that X_n is reversible.

Q2.6:

Find the stationary distribution of X_n denoted by $P\{X_n = i\} = \pi_i$.

Q2.7:

Compute the transition matrix, Q, for the time-reversed Markov chain, X_{-n}.

Solution:

Q2.1:

$$P = \begin{bmatrix} 1-\rho & \rho & 0 \\ 0 & 1-\rho & \rho \\ \rho & 0 & 1-\rho \end{bmatrix}$$

Q2.2:

In order to prove that it is irreducible, it is enough to show that for all state pairs, i, j, there exists and $n \geq 0$ such that $[P^n]_{i,j} > 0$.

Let $n = (j-i)\text{mod } 3$. Then $[P^n]_{i,j} > \rho^n > 0$. So the Markov chain is irreducible.

Q2.3:

Since $P_{0,0} > 1-\rho > 0$ and the Markov chain is irreducible by part Q2.2 above, then the Markov chain can not be periodic.

Q2.4:

Since the Markov chain has i) finite state; ii) is irreducible; iii) is aperiodic, then it must be ergodic.

Q2.5:

The Markov chain is not reversible. In order to prove this we need to show that

$$\pi_i P_{i,j} \neq \pi_j P_{j,i} \ ,$$

for some state pair, i, j.

By part Q2.6, we know that the stationary distribution is given by $\pi = [1/3, 1/3, 1/3]^t$. If we pick $i = 0$ and $j = 1$, then we have that $P_{0,1} = \rho$ and $P_{1,0} = 0$, so we have that

$$\pi_0 P_{0,1} = \rho/3 \neq 0 = \pi_1 P_{1,0} \ .$$

So since the detailed balance equations do not hold, the Markov chain is not reversible.

Q2.6:

Choose $\pi = [1/3, 1/3, 1/3]^t$. Then π_i satisfies the full balance equations given by

$$\pi P = \pi \ .$$

So it must be the stationary distribution of the ergodic Markov chain.

Q2.7:

We know that

$$\pi_i P_{i,j} = \pi_j Q_{j,i} \ .$$

So we have that

$$Q_{j,i} = P_{i,j} \frac{\pi_i}{\pi_j} = P_{i,j} \ .$$

So $Q = P^t$,

$$Q = \begin{bmatrix} 1-\rho & 0 & \rho \\ \rho & 1-\rho & \\ 0 & \rho & 1-\rho \end{bmatrix}$$

Q3: EM Algorithm (25pt)

Let $\{X_n\}_{n=1}^N$ be i.i.d. random variables with distribution

$$P\{X_n = m\} = \pi_m \ ,$$

where $\sum_{m=0}^{M-1} \pi_m = 1$. Also, let Y_n be conditionally independent random variables given X_n for $n = 1, \ldots, N$ with identical exponential conditional distributions given by

$$p(y_n|x_n = m) = u(y_n)\frac{1}{\mu_m} \exp\left\{-\frac{y_n}{\mu_m}\right\}$$

where

$$u(y) = \left\{ \begin{array}{l} 1 \text{ if } y \geq 0 \\ 0 \text{ if } y < 0 \end{array} \right. .$$

Furthermore, let $\theta = (\pi_0, \mu_0, \ldots, \pi_{M-1}, \mu_{M-1})$ parameterize the model.

Q3.1:

Write out the joint density function of $(X_1, Y_1, \ldots, X_N, Y_N)$ as a function of θ.

Q3.2:

Write out the joint density function for $(Y_1, \ldots, Y_N)$ as a function of θ.

Q3.3:

Write out an expression for the maximum likelihood estimate of θ given $(X_1, Y_1, \ldots, X_N, Y_N)$.

Q3.4:

Write out an expression for the E-step of the EM update of θ given $(Y_1, \ldots, Y_N)$. (Hint: Compute the posterior conditional expectation of the natural sufficient statistics of the distribution.)

Q3.5:

Write out an expression for the M-step of the EM update of θ given $(Y_1, \ldots, Y_N)$.

Solution:

Q3.1:

$$p(x,y) = \prod_{n=1}^{N} \left(\pi_{x_n} u(y_n) \frac{1}{\mu_{x_n}} \exp\left\{ -\frac{y_n}{\mu_{x_n}} \right\} \right) = \prod_{n=1}^{N} \frac{\pi_{x_n}}{\mu_{x_n}} \exp\left\{ -\frac{y_n}{\mu_{x_n}} \right\}$$

Q3.2:

$$p(y) = \prod_{n=1}^{N} \left(\sum_{m=0}^{M-1} \frac{\pi_m}{\mu_{x_n}} \exp\left\{ -\frac{y_n}{\mu_m} \right\} \right)$$

Q3.3:

Let

$$\begin{aligned} N_m &= \sum_{n=1}^{N} \delta(X_n - m) \\ S_m &= \sum_{n=1}^{N} Y_m \delta(X_n - m) \end{aligned}$$

Then

$$\begin{aligned} \hat{\pi}_m &= \frac{N_m}{N} \\ \hat{\mu}_m &= \frac{S_m}{N_m} \end{aligned}$$

Q3.4:

We can compute the posterior probability using Bayes' rule as

$$f(m|y_n, \theta) = P\{X_n = m | Y_n = y_n\} = \frac{\frac{\pi_m}{\mu_m} \exp\left\{ -\frac{y_n}{\mu_m} \right\}}{\sum_{k=0}^{M-1} \frac{\pi_k}{\mu_k} \exp\left\{ -\frac{y_n}{\mu_k} \right\}}$$

Then the E-step is given by

$$\begin{aligned} \bar{N}_m &= \sum_{n=1}^{N} f(m|y_n, \theta) \\ \bar{S}_m &= \sum_{n=1}^{N} Y_m f(m|y_n, \theta) \end{aligned}$$

Q3.5:

The M-step is given by

$$\begin{aligned}\hat{\pi}'_m &= \frac{\bar{N}_m}{N}\\ \hat{\mu}'_m &= \frac{\bar{S}_m}{\bar{N}_m}\end{aligned}$$

So then $\theta' = [\hat{\pi}'_m, \hat{\mu}'_m]$.

Q4: ADMM Optimization (20pt)

Consider the following MAP optimization problem

$$\hat{x} = \arg\min_{x} \{f(x) + h(x)\} \ ,$$

where $x \in \Re^N$.

Q4.1:

Write out the expression for the associated constrained optimization problem produced by splitting.

Q4.2:

Write out the associated augmented Lagrangian cost function $L(x, v; a, u)$.

Q4.3:

Write out pseudo-code for the augmented Lagrangian algorithm used for solving the MAP optimization problem.

Q4.4:

Write out pseudo-code for the ADMM algorithm for solving the MAP optimization problem.

Solution:

Q4.1:

$$\hat{x} = \arg \min_{(x,v) s.t. x=v} \{f(x) + h(v)\} \ ,$$

Q4.2:

$$L(x, v; a, u) = f(x) + h(v) + \frac{a}{2}\|x - v + u\|^2$$

Q4.3:

$u \leftarrow 0$

Init x, v

Repeat $\{$

$\quad (x, v) \leftarrow \arg\min_{x,v} \left\{f(x) + h(v) + \frac{a}{2}\|x - v + u\|^2\right\}$
$\quad u \leftarrow u + (x - v)$

$\}$

Q4.4:

$u \leftarrow 0$

Init x

Repeat $\{$

$\quad v \leftarrow \arg\min_{v} \left\{h(v) + \frac{a}{2}\|x - v + u\|^2\right\}$
$\quad x \leftarrow \arg\min_{x} \left\{f(x) + \frac{a}{2}\|x - v + u\|^2\right\}$
$\quad u \leftarrow u + (x - v)$

$\}$

Q5: Proximal Maps (25pt)

Consider the proximal maps given by

$$\hat{x} = F(z) = \arg\min_x \left\{ f(x) + \frac{1}{2\sigma^2}\|x - z\|^2 \right\}$$

$$\hat{x} = H(z) = \arg\min_x \left\{ \frac{1}{2\sigma^2}\|z - x\|^2 + h(x) \right\} \ ,$$

where f and h are continuously differentiable and convex.

Then our goal is to solve for the equilibrium conditions given by

$$F(x^* + u^*) = x^*$$

$$H(x^* - u^*) = x^* \ .$$

Q5.1:

Write out the basic plug-and-play ADMM algorithm for solving the equilibrium conditions.

Q5.2:

Prove that the values (x^*, u^*) that solve the equilibrium conditions also minimize the function $f(x) + h(x)$.

Q5.3:

Give an interpretation for the form of $H(z)$ as a MAP estimate. Provide an interpretation for each of the two terms in the cost function being minimized.

Q5.4:

Give an interpretation for the form of $F(z)$ as a MAP estimate. Provide an interpretation for each of the two terms in the cost function being minimized.

Q5.5:

If we were to replace $H(z)$ by a function that is trained using machine learning methods, then how should it be trained?

Solution:

Q5.1:

$u \leftarrow 0$

Init x

Repeat {

$v \leftarrow F(x+u)$

$x \leftarrow H(v-u)$

$u \leftarrow u + (x - v)$

}

Q5.2:

The equation $F(x^* + u^*) = x^*$ implies that

$$\nabla f(x^*) + \frac{1}{\sigma^2}(x^* - z)\Big|_{z=x^*+u^*} = \nabla f(x^*) - \frac{1}{\sigma^2}u^* = 0 \quad (1)$$

$$\nabla h(x^*) + \frac{1}{\sigma^2}(x^* - z)\Big|_{z=x^*-u^*} = \nabla h(x^*) + \frac{1}{\sigma^2}u^* = 0 \quad (2)$$

So therefore, we have that

$$\nabla \{f(x^*) + h(x^*)\} = 0 \ .$$

Since f and h and continuously differentiable and convex, x^* must be a global minimum.

Q5.3:

$$H(z) = \arg\min_x \left\{ \frac{1}{2\sigma^2}\|z - x\|^2 + h(x) \right\} \ ,$$

has the form of a MAP denoiser for the following virtual forward model of

$$Z = X + W \ ,$$

where $X \sim \frac{1}{z}\exp\{-h(x)\}$ is the prior model and $W \sim N(0, \sigma^2 I)$ is additive white Gaussian noise so that

$$-\log p(z|x) = \frac{1}{2\sigma^2}\|z - x\|^2 + \text{constant} \ ,$$

$$-\log p(x) = h(x) + \text{constant} \ .$$

Q5.4:

$$F(z) = \arg\min_x \left\{ f(x) + \frac{1}{2\sigma^2}\|x - z\|^2 \right\}$$

has the interpretation of solving for a MAP estimate where

$$-\log p(y|x) = f(x) \ ,$$

and

$$-\log p(x) = \frac{1}{2\sigma^2}\|x - z\|^2 \ .$$

So we are assuming that the prior distribution is $X \sim N(z, \sigma^2 I)$.

Q5.5:

$H(z)$ should be trained to remove additive white Gaussian noise with variance σ^2 from z. To do this, we can generate training pairs (Z_k, X_k) for $k = 0, \cdots, K-1$ so that

$$Z_k = X_k + W_k \ ,$$

where $W_k \sim N(0, \sigma^2 I)$. Then we can train the machine learning algorithm to minimize a loss with the form

$$Loss(\theta) = D(Z_k - H_\theta(X_k)) \ ,$$

where $D(\cdot)$ is a distortion function and θ parameterizes the ML function $f(\dot{)}$.