Final Exam - Moed A

Total time for the exam: 3 hours!

1) **Cross-Entropy of Three Predictors (42 Points):** A student trains a model to guess a label $X \in \mathcal{X} = \{1, 2, \ldots, M\}$ from a feature $Y \in \mathcal{Y}$. The training data and the test data come from one fixed joint distribution $P_{X,Y}$. A *predictor* is any conditional distribution $Q(x \mid y)$ on $\mathcal{X}$. Throughout the question, assume $P_{X,Y}(x, y) > 0$ and $Q(x \mid y) > 0$ for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$.
We give each predictor a score on the test data: its (conditional) **cross-entropy loss**

$$\mathcal{L}(Q) \triangleq H\big(P_{X,Y},\, Q_{X|Y}\big) \triangleq -\sum_{x,y} P_{X,Y}(x, y) \log Q(x \mid y) = -\,\mathbb{E}[\log Q(X \mid Y)]\,.$$

We look at three predictors:

$$Q_1(x \mid y) = 1/M, \qquad Q_2(x \mid y) = P_X(x), \qquad Q_3(x \mid y) = P_{X|Y}(x \mid y).$$

All logarithms are base 2.

a) **(4 points)** Find $\mathcal{L}(Q_1)$.
Solution:

$$\mathcal{L}(Q_1) = -\sum_{x,y} P_{X,Y}(x, y) \log \tfrac{1}{M} = \log M \cdot \sum_{x,y} P_{X,Y}(x, y) = \boxed{\log M}.$$

b) **(4 points)** Find $\mathcal{L}(Q_2)$.
Solution:

$$\mathcal{L}(Q_2) = -\sum_{x,y} P_{X,Y}(x, y) \log P_X(x) = -\sum_{x} P_X(x) \log P_X(x) = \boxed{H(X)}.$$

(The sum over y marginalises out; no joint information is needed.)

c) **(4 points)** Find $\mathcal{L}(Q_3)$.
Solution:

$$\mathcal{L}(Q_3) = -\sum_{x,y} P_{X,Y}(x, y) \log P_{X|Y}(x \mid y) = \mathbb{E}_Y\Big[-\sum_{x} P(x \mid Y) \log P(x \mid Y)\Big] = \boxed{H(X \mid Y)},$$

by the definition $H(X \mid Y) = \mathbb{E}_Y[H(X \mid Y{=}y)]$.

d) **(4 points) True or False:** for every joint $P_{X,Y}$, $\mathcal{L}(Q_1) \geq \mathcal{L}(Q_2) \geq \mathcal{L}(Q_3)$. Explain in at most 2 lines.
Solution: True. Using parts (a)–(c) and standard Lecture-1–2 facts:

$$\mathcal{L}(Q_1) - \mathcal{L}(Q_2) = \log M - H(X) \;\geq\; 0 \qquad \text{(entropy is largest for the uniform distribution)},$$
$$\mathcal{L}(Q_2) - \mathcal{L}(Q_3) = H(X) - H(X \mid Y) = I(X;Y) \;\geq\; 0 \qquad \text{(conditioning reduces entropy)}.$$

e) **(4 points)** For each equality below, give the requested condition.

(i) **(2 points)** $\mathcal{L}(Q_1) = \mathcal{L}(Q_2)$. Condition on P_X.
Solution: $\mathcal{L}(Q_1) = \mathcal{L}(Q_2) \iff \log M = H(X) \iff \boxed{P_X \text{ is uniform on } \mathcal{X}}$.

(ii) **(2 points)** $\mathcal{L}(Q_3) = 0$. Condition on $P_{X,Y}$.
Solution: $\mathcal{L}(Q_3) = 0 \iff H(X \mid Y) = 0 \iff \boxed{X \text{ is a deterministic function of } Y}$.

f) **(6 points)** With $V = g(Y)$ and $Q_4(x \mid y) \triangleq P_{X|V}(x \mid g(y))$, show that $\mathcal{L}(Q_4) - \mathcal{L}(Q_3) = I(X; Y \mid V)$.
Solution: Since $Q_4(X \mid Y) = P_{X|V}(X \mid g(Y)) = P_{X|V}(X \mid V)$,

$$\mathcal{L}(Q_4) = -\mathbb{E}\big[\log P_{X|V}(X \mid V)\big] = H(X \mid V).$$

Because $V = g(Y)$ is a deterministic function of Y, knowing Y already fixes V, so $H(X \mid Y, V) = H(X \mid Y)$. Therefore

$$\mathcal{L}(Q_4) - \mathcal{L}(Q_3) = H(X \mid V) - H(X \mid Y) = H(X \mid V) - H(X \mid Y, V) = \boxed{I(X; Y \mid V)},$$

the conditional mutual information between X and Y given V: the information about X that the full feature Y still carries beyond the simpler view $V = g(Y)$.

g) **(4 points)** Find $\mathbb{E}[Z]$ for $Z \triangleq Q(X \mid Y)/P_{X|Y}(X \mid Y)$.
Solution: Use the chain rule $P_{X,Y}(x, y) = P_Y(y)\, P_{X|Y}(x \mid y)$:

$$\mathbb{E}[Z] = \sum_{x,y} P_{X,Y}(x, y) \frac{Q(x \mid y)}{P_{X|Y}(x \mid y)} = \sum_{y} P_Y(y) \underbrace{\sum_{x} Q(x \mid y)}_{=1} = \boxed{1}.$$

h) **(12 points) True or False:** for every predictor Q, $\mathcal{L}(Q) \geq \mathcal{L}(Q_3)$. If true, provide a proof; if false, provide a counterexample.

Solution: True. Two proofs are given below; either earns full credit.

Option 1 — Jensen's inequality (via part (g)). Since $-\log$ is convex, Jensen's inequality applied to the random variable Z from part (g) gives

$$\mathbb{E}[-\log Z] \;\geq\; -\log \mathbb{E}[Z] = -\log 1 = 0.$$

Now expand $-\log Z = -\log Q(X \mid Y) + \log P_{X|Y}(X \mid Y)$ and take expectation under $P_{X,Y}$:

$$\mathbb{E}[-\log Z] = \mathcal{L}(Q) - \mathcal{L}(Q_3) \;\geq\; 0, \qquad \text{i.e.} \qquad \mathcal{L}(Q) \geq \mathcal{L}(Q_3). \qquad \blacksquare$$

Option 2 — KL divergence Write the gap directly as an expected log-ratio and regroup by y:

$$\begin{aligned}\mathcal{L}(Q) - \mathcal{L}(Q_3) &= \sum_{x,y} P_{X,Y}(x,y) \log \frac{P_{X|Y}(x \mid y)}{Q(x \mid y)} = \sum_y P_Y(y) \sum_x P_{X|Y}(x \mid y) \log \frac{P_{X|Y}(x \mid y)}{Q(x \mid y)} \\ &= \sum_y P_Y(y)\, D\big(P_{X|Y}(\cdot \mid y) \,\|\, Q(\cdot \mid y)\big) \;=\; D\big(P_{X|Y} \,\|\, Q_{X|Y} \mid P_Y\big) \;\geq\; 0,\end{aligned}$$

since every KL divergence is non-negative. Hence $\mathcal{L}(Q) \geq \mathcal{L}(Q_3)$.

$$\text{Equality holds} \iff Q(\cdot \mid y) = P_{X|Y}(\cdot \mid y) \text{ for every } y \text{ (i.e. } Q = Q_3). \qquad \blacksquare$$

2) **Binary Channel with Two Noisy Observations (42 Points):**
Consider a memoryless channel with binary input $X \in \{0,1\}$ and output $Y = (Y_1, Y_2) \in \{0,1\}^2$. For each channel use,

$$Y_1 = X \oplus Z_1, \qquad Y_2 = X \oplus Z_2,$$

where $\oplus$ denotes addition modulo 2, and

$$Z_1, Z_2 \overset{\text{i.i.d.}}{\sim} \text{Bern}(p), \qquad 0 \leq p \leq \frac{1}{2},$$

independent of X.

a) **(8 points)** Characterize the channel as follows.

(i) **(4 points)** Write the channel transition matrix

$$W(y_1, y_2 \mid x) = \Pr(Y_1 = y_1, Y_2 = y_2 \mid X = x).$$

Solution:
Since $Y_j = X \oplus Z_j$ and Z_1, Z_2 are independent, the channel transition matrix is

$$W = \begin{array}{c|cccc} & (0,0) & (0,1) & (1,0) & (1,1) \\ \hline X=0 & (1-p)^2 & p(1-p) & p(1-p) & p^2 \\ X=1 & p^2 & p(1-p) & p(1-p) & (1-p)^2 \end{array}.$$

(ii) **(4 points)** Write the formula for the capacity of a general channel with input X and output Y. Then write the corresponding formula for the channel above, whose output is (Y_1, Y_2), and denote its capacity by $C_2(p)$.
Solution:
The capacity of a general channel with input X and output Y is

$$C = \max_{P_X} I(X;Y).$$

Therefore, the capacity of the channel above is

$$\boxed{C_2(p) = \max_{P_X} I(X; Y_1, Y_2)}.$$

b) **(6 points)** Find a capacity-achieving input distribution for this channel.
Hint: Examine the symmetry between the two rows of the transition matrix.
Solution:
The row corresponding to $X = 1$ is obtained from the row corresponding to $X = 0$ by the permutation

$$(0,0) \longleftrightarrow (1,1), \qquad (0,1) \longleftrightarrow (1,0).$$

Thus, the channel is symmetric. Therefore, the uniform input distribution is capacity-achieving:

$$\boxed{P_X(0) = P_X(1) = \frac{1}{2}}.$$

c) **(12 points)** Use the capacity-achieving input distribution found in the previous part.

(i) **(4 points)** Compute $H(Y_1, Y_2 \mid X)$.

Solution:
Given X, the pair (Y_1, Y_2) is obtained from the independent noise variables (Z_1, Z_2). Therefore,

$$H(Y_1, Y_2 \mid X) = H(Z_1, Z_2).$$

Since Z_1 and Z_2 are independent,

$$H(Z_1, Z_2) = H(Z_1) + H(Z_2) = 2h_2(p),$$

where $h_2(\cdot)$ is the binary entropy function. Hence,

$$\boxed{H(Y_1, Y_2 \mid X) = 2h_2(p)}.$$

(ii) **(5 points)** Find the probability distribution of (Y_1, Y_2) and compute

$$H(Y_1, Y_2).$$

Solution:
Using $P_X(0) = P_X(1) = \frac{1}{2}$,

$$P(0,0) = P(1,1) = \frac{(1-p)^2 + p^2}{2},$$

and

$$P(0,1) = P(1,0) = p(1-p).$$

Hence,

$$P_{Y_1,Y_2} = \left[\frac{1-2p(1-p)}{2},\ p(1-p),\ p(1-p),\ \frac{1-2p(1-p)}{2}\right].$$

Let $q = 2p(1-p)$. Then

$$P_{Y_1,Y_2} = \left[\frac{1-q}{2}, \frac{q}{2}, \frac{q}{2}, \frac{1-q}{2}\right],$$

and therefore

$$\boxed{H(Y_1, Y_2) = 1 + h_2\big(2p(1-p)\big)}.$$

(iii) **(3 points)** Compute $C_2(p)$ as a function of p.
Solution:
The uniform input distribution is capacity-achieving. Therefore,

$$\begin{aligned} C_2(p) &= I(X; Y_1, Y_2) \\ &= H(Y_1, Y_2) - H(Y_1, Y_2 \mid X) \\ &= 1 + h_2\big(2p(1-p)\big) - 2h_2(p). \end{aligned}$$

Hence,

$$\boxed{C_2(p) = 1 + h_2\big(2p(1-p)\big) - 2h_2(p)}.$$

d) **(8 points)** Assume that X is uniformly distributed over $\{0, 1\}$.

(i) **(4 points)** Determine the MAP estimate $\widehat{X}(y_1, y_2)$ for each of the four possible output pairs. If the two possible input values have equal posterior probabilities, choose between them uniformly at random.
Solution:
Since the input distribution is uniform, the MAP rule compares

$$W(y_1, y_2 \mid 0) \quad \text{and} \quad W(y_1, y_2 \mid 1).$$

For $0 \le p < \frac{1}{2}$, the decisions are

(y_1, y_2)	$W(y_1, y_2 \mid 0)$	$W(y_1, y_2 \mid 1)$	$\widehat{X}(y_1, y_2)$
$(0,0)$	$(1-p)^2$	p^2	0
$(0,1)$	$p(1-p)$	$p(1-p)$	0 or 1 uniformly
$(1,0)$	$p(1-p)$	$p(1-p)$	0 or 1 uniformly
$(1,1)$	p^2	$(1-p)^2$	1

Thus, when the two observations agree, the receiver chooses their common value. When they disagree, the receiver chooses uniformly at random.
At $p = \frac{1}{2}$, both input values have equal posterior probability for every output pair, so the receiver chooses uniformly at random for all four outputs.

(ii) **(4 points)** Compute the resulting probability of error

$$P_e = \Pr(\widehat{X} \neq X).$$

Solution:
An error occurs in one of the following cases:

- Both observations are incorrect, which has probability p^2.
- Exactly one observation is incorrect. This has probability $2p(1-p)$, and the random tie-breaking decision is incorrect with probability $\frac{1}{2}$.

Therefore,

$$\begin{aligned} P_e &= p^2 + \frac{1}{2}\big(2p(1-p)\big) \\ &= p^2 + p(1-p) \\ &= p. \end{aligned}$$

Thus, two observations improve the channel capacity, but with only two observations and random tie breaking, they do not reduce the MAP error probability compared with observing only one BSC output.

e) **(8 points)** Answer the following questions about processing the channel output.

(i) **(3 points)** Suppose that instead of observing the complete pair (Y_1, Y_2), the receiver observes only

$$T = Y_1 \oplus Y_2.$$

Compute the capacity of the induced channel from X to T.
Solution:
We have

$$\begin{aligned} T &= Y_1 \oplus Y_2 \\ &= (X \oplus Z_1) \oplus (X \oplus Z_2) \\ &= Z_1 \oplus Z_2. \end{aligned}$$

Thus, T does not depend on X. Accordingly, it follows that $I(X;T) = 0$ for every input distribution. Hence,

$$\boxed{C = 0}.$$

(ii) **(3 points)** A receiver that observes only Y_1 sees a binary symmetric channel with crossover probability p, whose capacity is $C_1(p) = 1 - h_2(p)$. Compare $C_1(p)$ with the capacity $C_2(p)$ found in part (c).
Solution:
Using the expressions for the two capacities,

$$\begin{aligned} C_2(p) - C_1(p) &= \big[1 + h_2\big(2p(1-p)\big) - 2h_2(p)\big] - [1 - h_2(p)] \\ &= h_2\big(2p(1-p)\big) - h_2(p). \end{aligned}$$

For

$$0 < p < \frac{1}{2},$$

we have

$$p < 2p(1-p) < \frac{1}{2}.$$

Since $h_2(q)$ is strictly increasing for $0 \leq q \leq \frac{1}{2}$,

$$h_2\big(2p(1-p)\big) \geq h_2(p).$$

Therefore,

$$C_2(p) \geq C_1(p), \qquad 0 < p < \frac{1}{2}.$$

(iii) **(2 points)** Evaluate $C_2(p)$ at $p = 0$ and at $p = \frac{1}{2}$, and briefly explain the operational meaning of each result.
Solution:
For $p = 0$,

$$C_2(0) = 1 + h_2(0) - 2h_2(0) = 1.$$

Thus,

$$\boxed{C_2(0) = 1 \text{ bit per channel use}}.$$

In this case, there is no noise and

$$Y_1 = Y_2 = X,$$

so the input bit is recovered perfectly.

For $p = \frac{1}{2}$,

$$C_2\left(\frac{1}{2}\right) = 1 + h_2\left(\frac{1}{2}\right) - 2h_2\left(\frac{1}{2}\right) = 1 + 1 - 2 = 0.$$

Thus,

$$\boxed{C_2\left(\frac{1}{2}\right) = 0}.$$

In this case, each output is an independent fair bit and contains no information about X, even when both observations are available.

‘‘‘

3) **Convolutional Codes and Viterbi Decoding (41 Points):**
Consider a binary convolutional encoder with one input bit u_t at each time and two output bits $v_t = (v_t^{(1)}, v_t^{(2)})$. The generator sequences are

$$g^{(1)} = (1, 0, 1), \qquad g^{(2)} = (1, 1, 1).$$

Let $\mathbf{u} = (u_1, u_2, \ldots)$. For generator $g^{(i)}$, define $v_t^{(i)} = (\mathbf{u} * g^{(i)})_t = \sum_{\ell=0}^{m} g_\ell^{(i)} u_{t-\ell}$, where $*$ is convolution over $\mathbb{F}_2$. The encoder starts in the all-zero state.

a) **(5 points)** Draw the encoder diagram. Define the convolutional code parameters (n, k, m), its rate, and the number of states. Define the state s_t and the output vector v_t.

b) **(5 points)** Show the state transitions, either as a table or as a state diagram. Each branch should be labeled by

$$\text{input/output}.$$

c) **(5 points)** We want to transmit the information bits

$$(1, 1, 0)$$

using the convolutional code above. Assume the decoder knows that the final state is 00. Write the serialized transmit vector

$$V = (v_1^{(1)}, v_1^{(2)}, v_2^{(1)}, v_2^{(2)}, \ldots).$$

What is the length of V?

d) **(18 points)** For this part, the convolutional encoder above is used to transmit two information bits. The resulting coded bits are mapped to BPSK symbols by

$$0 \mapsto +1, \qquad 1 \mapsto -1.$$

The symbols are transmitted over an AWGN channel

$$Y_i = X_i + N_i, \qquad N_i \sim \mathcal{N}(0, \sigma^2),$$

with $\sigma^2 = 1$. The receiver observes

$$y = (1, 1,\ 1, -2,\ -2, 1,\ 1, 1).$$

The squared Euclidean metric is

$$d_E^2(x, y) = \sum_i (y_i - x_i)^2.$$

Assume the trellis path starts and ends in state 00. **Explain your answer in each part.**

(i) **(6 points)** Apply hard decisions to the received samples, and then use Viterbi decoding with Hamming-distance branch metrics. What are the estimated transmitted information bits? Explain your answer.

(ii) **(8 points)** Apply Viterbi decoding using the squared Euclidean metric. What are the estimated transmitted information bits? Explain your answer.

(iii) **(4 points)** Compare Viterbi decoding with the Hamming metric and with the squared Euclidean metric. Which metric is better for this received vector, and why can the two decoders give different estimated information bits? Explain your answer.

e) **(8 points)** Answer briefly:

(i) What is the difference between what the Viterbi algorithm outputs and what the BCJR algorithm computes?

(ii) What is the decoding algorithm used in a turbo-code decoder: BCJR, Viterbi as in part (d), or can both be used? Explain why in one or two lines.

Solution:

a) At each time the encoder receives $k = 1$ input bit and outputs $n = 2$ coded bits. Since the longest generator has length 3, the memory is $m = 2$. Therefore this is an $(n, k, m) = (2, 1, 2)$ convolutional code, its rate is

$$R = \frac{k}{n} = \frac{1}{2},$$

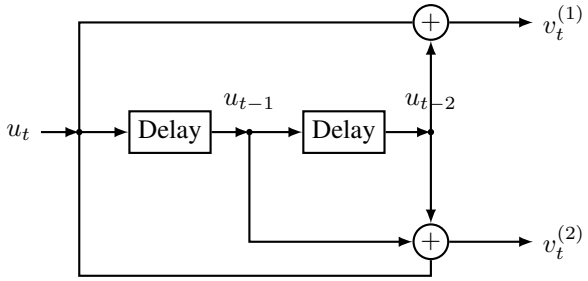
and it has $2^m = 4$ states. A natural state is

$$s_t = (u_{t-1}, u_{t-2}).$$

The output vector is

$$v_t = (v_t^{(1)}, v_t^{(2)}) = (u_t \oplus u_{t-2},\ u_t \oplus u_{t-1} \oplus u_{t-2}).$$

The encoder has rate $R = k/n = 1/2$ and $2^m = 4$ states. The encoder diagram has two delay elements storing u_{t-1} and u_{t-2}; the first XOR gate outputs $u_t \oplus u_{t-2}$, and the second XOR gate outputs $u_t \oplus u_{t-1} \oplus u_{t-2}$. One possible diagram is:



b) The complete transition table is:

u_t	current state	next state	output
0	00	00	00
1	00	10	11
0	01	00	11
1	01	10	00
0	10	01	01
1	10	11	10
0	11	01	10
1	11	11	01

c) Since the encoder memory is $m = 2$, we add two zero tail bits. The full encoder input is

$$(1, 1, 0, 0, 0).$$

Starting from state 00, the state path is

$$00 \to 10 \to 11 \to 01 \to 00 \to 00.$$

Therefore the end state is 00. The output pairs are

$$(11, 10, 10, 11, 00).$$

Therefore the serialized transmit vector is

$$V = (1, 1, 1, 0, 1, 0, 1, 1, 0, 0).$$

Here the number of information bits is $L = 3$. After termination, the trellis has

$$L + m = 3 + 2 = 5$$

stages. Since each stage produces $n = 2$ coded bits, the length of the serialized transmit vector is

$$|V| = n(L + m) = 2(3 + 2) = 10.$$

d) The received vector contains four trellis sections. Since two information bits are transmitted and the tail length is $m = 2$, each valid input sequence has two information bits followed by two tail bits:

$$(u_1, u_2, 0, 0).$$

The hard decisions from the received AWGN samples are

$$r = (00, 01, 10, 00).$$

The four possible valid paths are:

information bits	full input sequence	coded output pairs
$(0, 0)$	$(0, 0, 0, 0)$	$(00, 00, 00, 00)$
$(0, 1)$	$(0, 1, 0, 0)$	$(00, 11, 01, 11)$
$(1, 0)$	$(1, 0, 0, 0)$	$(11, 01, 11, 00)$
$(1, 1)$	$(1, 1, 0, 0)$	$(11, 10, 10, 11)$

(i) For hard-decision Viterbi decoding, the branch metric is the Hamming distance between each received pair and each

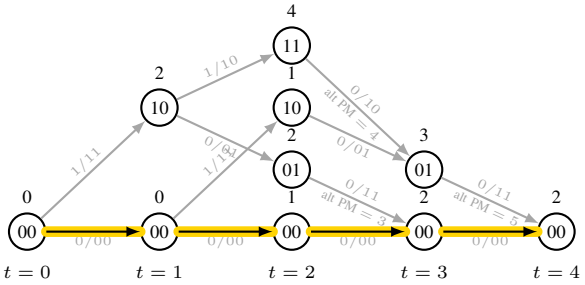
branch output. The total Hamming distances of the valid paths are:

information bits	coded output pairs	$d_H(r, v)$
$(0, 0)$	$(00, 00, 00, 00)$	2
$(0, 1)$	$(00, 11, 01, 11)$	5
$(1, 0)$	$(11, 01, 11, 00)$	3
$(1, 1)$	$(11, 10, 10, 11)$	6

Thus hard-decision decoding chooses the path corresponding to information bits $(0, 0)$. Equivalently, the Viterbi path metrics are shown below. Since the final state is known to be 00, at $t = 3$ only states 00 and 01 can still reach 00 in one remaining transition; states 10 and 11 are therefore invalid.

t	$PM_t(00)$	$PM_t(01)$	$PM_t(10)$	$PM_t(11)$
0	0	∞	∞	∞
1	0	∞	2	∞
2	1	2	1	4
3	2	3	invalid	invalid
4	2	invalid	invalid	invalid

The hard-decision Viterbi recursion is shown graphically below. The state is written inside each node, and the path metric is written above the node. The highlighted path is the selected hard-decision survivor path.



Backtracking from state 00 gives

$$\hat{u}_{\text{hard}} = (0, 0, 0, 0),$$

so the hard-decision estimate of the information bits is

$$\hat{u}_{\text{hard,info}} = (0, 0).$$

(ii) For Viterbi decoding with the squared Euclidean metric, the branch metric is

$$d_E^2(y_t, x_t) = \sum_{j=1}^{2} \left(y_{t,j} - (1 - 2v_t^{(j)}) \right)^2 .$$

For example, at $t = 1$, $y_1 = (1, 1)$, so

$$d_E^2(y_1, 00) = (1 - 1)^2 + (1 - 1)^2 = 0,$$

while

$$d_E^2(y_1, 11) = (1 + 1)^2 + (1 + 1)^2 = 8.$$

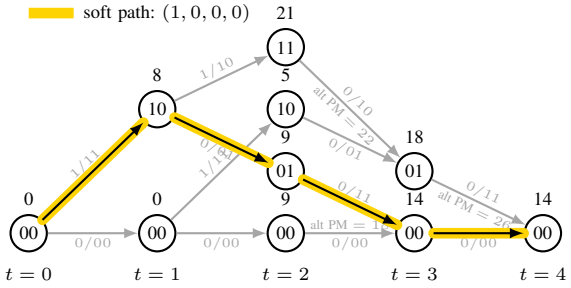
The resulting accumulated squared Euclidean branch metrics for the valid paths are:

information bits	coded output pairs	$\sum_t d_E^2(y_t, x_t)$
$(0, 0)$	$(00, 00, 00, 00)$	18
$(0, 1)$	$(00, 11, 01, 11)$	26
$(1, 0)$	$(11, 01, 11, 00)$	14
$(1, 1)$	$(11, 10, 10, 11)$	30

Therefore the minimum squared Euclidean metric is obtained by the path corresponding to information bits $(1, 0)$. Equivalently, using accumulated d_E^2 branch metrics inside the Viterbi recursion gives the table below. Again, states 10 and 11 are invalid at $t = 3$ because they cannot reach final state 00 in one remaining transition.

t	$PM_t(00)$	$PM_t(01)$	$PM_t(10)$	$PM_t(11)$
0	0	∞	∞	∞
1	0	∞	8	∞
2	9	9	5	21
3	14	18	invalid	invalid
4	14	invalid	invalid	invalid

The soft-decision Viterbi recursion is shown graphically below. The state is written inside each node, and the path metric is written above the node. The highlighted path is the selected soft-decision survivor path.



Backtracking from state 00 gives

$$\hat{u}_{\text{soft}} = (1, 0, 0, 0),$$

so the soft-decision estimate of the information bits is

$$\hat{u}_{\text{soft,info}} = (1, 0).$$

(iii) For this received vector, the squared Euclidean metric is better because it is equivalent to the maximum-likelihood metric for the AWGN channel. The selected squared-Euclidean-metric path is

$$(11, 01, 11, 00),$$

while Hamming-metric decoding chooses the all-zero path. The difference occurs because the Hamming metric uses only hard decisions, so it keeps only the sign of each sample. The samples -2 are stronger evidence for coded bit 1 than samples close to 0 would be, and the squared Euclidean metric keeps this magnitude information. Therefore it can prefer the path that better matches the more reliable samples.

e) (i) Viterbi finds the most likely complete path, or equivalently the most likely transmitted sequence. BCJR computes symbol-by-symbol posterior probabilities such as $P(u_t = 0 \mid y)$ and $P(u_t = 1 \mid y)$, often represented by an APP L-value.

(ii) Turbo-code decoders usually use BCJR, in MAP or log-MAP form. A soft-output Viterbi algorithm can also be used as an approximation, but the ordinary Viterbi algorithm from part (d) is not enough because turbo decoding needs soft APP/extrinsic LLRs to pass between the component decoders.

Good Luck!