This is homework 1 for "Understanding Vision 1 (lecture), Summer 2024", handed out June 9, due June 21, 2024

Please answer the following multiple choice questions. **These choices are marked as A, B, C, etc. To pick a choice, e.g., choice A, please draw a circle around this letter, e.g., A.** For each question, pick one or more choices that you think are correct.

1. if a random integer number can take at most 8 possible values: 0, 1, 2, 3, 4, 5, 6, 7, what is the minimum number of yes-no questions you need to ask in order to definitely nail it down for any of these 8 possibilities? Note that this number is the number of bits that is needed to encode this integer when this integer can take any of these 8 values with equal probability.

   A. 3

   B. 8

   C. 1

2. For a random number that is a single digit binary number, which can take value 0 or 1 with equal probability, what is the entropy of this random number?

   A. 1 bit.

   B. 2 bit

   C. 0.5 bit

3. if O = S+N, and S is an integer, and N is an integer that can take three possible values N = 0, N=1, N=-1 equally likely, then, if we observe that O = 5, what is P(S—O) for S=4? In other words, what is the conditional probability that of S = 4 given that O = 5?

   A. 1/3

   B. 1

   C. 0

4. if S can take only two possible values S = 0 and S = 1 with equal probability, if O can take also only two possible values O = 0 and O = 1 with equal probability, and O = S in 90% of the cases, and O and S are not equal to each other due to noise in 10% of the cases, what is the mutual information between O and S?

   A. 1 bit

   B. 0.531 bit

5. If your signal S is a gaussian variable with mean 0 and standard deviation = 1, and your noise N is a gaussian variable with mean 0 and standard deviation = 1, what is O = S+N?

   A. a gaussian with mean zero and standard deviation 1.414 (square root of 2)

   B. a gaussian with mean 0 and standard deviation = 2

6. Imagine you have an image, in which each pixel takes one of the 256 integer values, S_1 = 0, 1, 2, 3, ..., 255 with equal probability, so that the entropy of this pixel value is $log_2(256) = 8$ bits. Now you make a binary version of this image, so that each pixel takes only 2 values, $S_2 = 0$ or $S_2 = 1$. S_2 for a pixel in the binary image is uniquely determined by its pixel value S_1 of this same image pixel in the original image as follows, S_2 = 0, if $S_1 < 128$; $S_2 = 1$, if $S_1 \geq 128$. We can see that S_2 takes the value of 0 or 1 with equal probability.

   So we have entrory $H(S_1) = 8$bits, $H(S_2) = 1$ bit (since S2 = 0 or 1 with equal probability). The redundancy between S_1 and S_2 is

   A. 1/8

   B. 1

7. Efficient coding is formulated as follows: to find an optimal encoding transform K for visual input signal $\mathbf{S}$, to get the encoded outcome $\mathbf{O}$, such that this objective function value $E(K) =$ neural cost $-\lambda \cdot I(O; S)$ is minimized. In the expression for $E(K)$, $I(\mathbf{O};\mathbf{S})$ is the mutual information between $\mathbf{O}$ and $\mathbf{S}$, and λ is a parameter to balance between the benefit $I(\mathbf{O};\mathbf{S})$ and the neural cost. If it is important to save cost, then you should

   A. reduce the value of λ in $E(K)$.

   B. increase the value of λ in $E(K)$

8. Efficient in the zero input noise limit is related to, or equivalent to,

   A. maximum entropy coding

   B. Independent components coding

   C. factorial codes

   D. decorrelation

   E. redundancy reduction

9. In efficient coding in the high input noise limit, it is useful to

   A. average over redundant input signal channels to smooth out the noise

   B. average over the redundant input signal channels to preserve the input redundancy

10. Which one is a reasonable choice for neural cost

   A. total output channel capacity

   B. total output power or dynamic range

   C. total input power

11. To describe the probability distribution $P(\mathbf{S})$ of the visual input signals $\mathbf{S}$, we approximate $P(\mathbf{S})$ as Gaussian. What are the reasons for using this approximation?

   A. Because $\mathbf{S}$ is multi-dimensional, so we often do not know what the true $P(\mathbf{S})$ is.

   B. Gaussian $P(\mathbf{S})$ captures our ignorance about the true $P(\mathbf{S})$, because it has the maximum entropy given the correlation matrix R between different components of $\mathbf{S}$.

   C. It makes it easier to solve for an analytical solution for the efficient encoding matrix K.

   D. Gaussian approximation is not too bad an approximation.

12. What are the three concept components in the efficient encoding transform K (formulated as a matrix for a linear encoding of gaussian input signal $\mathbf{S}$ by output $\mathbf{O}$) for receptive fields?

   A. principal component analysis, gain control, and multiplexing in this order.

   B. principal component analysis, multiplexing , and gain control in this order

13. What is the reason behind the third component, the multiplexing matrix U (which is orthonormal or unitary) in the solution K for the efficient coding transform to minimize the objective function $E(\mathsf{K})$?

   A. because there are many possible solutions for the most efficient transform K, and that these solutions are equally efficient to give the same objective function value $E(\mathsf{K})$.

   B. Another way to state the choice (1) above is that the solutions for the most efficient transform K is degenerate, captured by this multiplexing matrix U.

   C. because there are many disjoint local minimums in the objective function $E(\mathsf{K})$, each minimum gives a solution K, the matrix transform U transitions discontinuously from one local minimum to another.

14. what does the principal component transform do to the input signal vector $\mathbf{S}$?

    A. it rotates the coordinate system for $\mathbf{S}$, so that in this new coordinate system, the components for $\mathbf{S}$ are independent of each other.

    B. It does a Fourier transform on the $\mathbf{S}$ vector.

15. In the efficient coding matrix K, what determines the gain value, g_k, for each principal component k?

    A. The gain value should be the one to minimize the whole objective function $E(\mathsf{K})$, by minimizing for each component k.

    B. The gain value is determined to minimize the total output cost.

16. In the efficient linear coding matrix K, the gain value g_k, for each principal component of $\mathbf{S}$, depends on

    A. the input signal to noise of this principal component

    B. the λ value in the objective function $E(\mathsf{K})$.

17. Once we have the solution K for the efficient transform from the photoreceptors $\mathbf{S}$ to the encoding responses $\mathbf{O}$ of the retinal ganglion cells, could we build an artificialr network with two layers, one for the input layer for vector $\mathbf{S}$ and one for the output layer for the encoding vector $\mathbf{O}$, and K is used for the connections from $\mathbf{S}$ to $\mathbf{O}$, and use this artificial network to substitute for the retina?

    A. yes, if all one wants is the efficient coding for a particular situation..

    B. Yes, as above, but when the environment changes, e.g., probability $P(\mathbf{S})$ may change, then the optimal solution K should change, then we have to change this artificial network.

18. In this example of ocularity coding for inputs from two eyes, what is the dimension for input signal $\mathbf{S}$?

    A. 1

    B. 2, S_L for the left eye, S_R for the right eye.

19. In the formulation of binocular inputs $\mathbf{S}$, where are the encoding neurons for encoding responses O?

    A. in the retina, as the retinal ganglion cells.

    B. in V1, the primary visual cortex.

20. For encoding binocular inputs $\mathbf{S}$ by responses $\mathbf{O}$, ignoring other feature dimensions, what are the dimensions of the encoding transform K?

    A. 2×2

    B. $N \times N$

21. Please tick all answers that apply below regarding the correlation matrix R^S for the two dimensional signal $\mathbf{S}$ of the binocular inputs.

    A. It is a 2×2 symmetric matrix

    B. In normal cases (when the two eyes are statistically the same in terms of inputs), the diagonal elements are expected to be identical to each other.

22. Following questions above for encoding the binocular inputs $\mathbf{S}$, how many principal components are there for $\mathbf{S}$?

    A. 2, same as the dimension of $\mathbf{S}$.

    B. N.

23. Which principal component has a stronger variance in normal situations, when binocular inputs for left and right eyes $\mathbf{S} = (S_L, S_R)$, and S_L and S_R have equal variance?

    A. the ocular summation one, representing $S_L + S_R$

    B. the ocular difference one, representing $S_L - S_R$.

24. For encoding the 2-d binocular inputs $\mathbf{S}$, the principal component transform K_o, in the normal case when the two eyes are equal in their signal variance, is

A. a 45 degree rotation in the 2-d space for $\mathbf{S}$.

B. a rotation and a scaling in 2-d $\mathbf{S}$ space.

25. The two principal components for binocular inputs $\mathbf{S}$ are denoted as S_+ and S_-, are they correlated with each other?

A. no

B. yes

26. if $\mathbf{O}$ is the outcome from the principal component transform of S (and with no gain control or multiplexing),

what is the correlation matrix for output vector $\mathbf{O}$ like, in normal situations when the inputs to the two eyes, S_L and S_R, are correlated with each other?

A. it is a diagonal matrix, with unequal diagonal elements, one for the ocular summation component, and one for the ocular difference component

B. It is not a diagonal matrix, and its details depend.

27. In encoding the two dimensional binocular input signal $\mathbf{S}$, how many gains need to be determined by the efficient coding transform?

A. two gains, one for the ocular summation component S_+, another for the ocular difference component S_-,

B. four gains, for the left eye input S_L, for the right eye input S_R, for the summation S_+, and for the difference S_-.

28. About the two gains, g_+ and g_-, for the ocular summation and ocular difference components of binocular input signal $\mathbf{S}$, please tick all statements below that are correct for the situation when the input noise is very low

A. $g_+ < g_-$

B. $g_+ > g_-$

C. we need a higher gain g_- to enhance the ocular contrast, which has the information regarding stereo depth.

29. About the two gains, g_+ and g_-, for the ocular summation and ocular difference components of binocular input signal $\mathbf{S}$, please tick all statements below that are correct for the situation when the input noise is much higher than input signal.

A. $g_+ < g_-$

B. $g_+ > g_-$

C. g_- should be low to avoid passing too much input noise to the output encoding responses.

30. For the efficient linear encoding of the 2-dimensional binocular input signals $\mathbf{S}$, what could be the U matrix, the third step in the encoding matrix K, for the multiplexing?

A. another 45 degree rotation in 2-d space.

B. could be just an identity 2×2 matrix.

31. Following from above, one good U matrix is the one that is a coordinate rotation by -45^o, the inverse of the principal component transform step K_o. Please tick all statements that are correct below.

A. This U matrix is special, so that, if there is no gain control (so that $\mathsf{g} = 1$), then the whole efficient coding transform matrix $\mathsf{K} = \mathsf{K}_o\mathsf{g}\mathsf{U}$ would be just an identical matrix, with no transform of the input signal $\mathbf{S}$.

B. It mixes the two gain controlled components, $g_+\mathcal{S}_+$ and $g_-\mathcal{S}_-$ (where $\mathcal{S}_+$ and $\mathcal{S}_-$ are the two principal components). After this mixing, each output node, O_1 or O_2, has typically unequally weighted mixture of the original left eye S_L and right eye S_R signals, so to be dominated more or less by inputs from one eye over another, when, in general cases, g_+ and g_- have different values.

32. In the bioncular encoding, when input environment becomes dimmer, so that signal to noise decreases, how does the encoding adapt?

    A. g_+/g_- increases, to devote the encoding more to the binocular redundant component and less to the ocular contrast which contains too much noise.

    B. g_+/g_- decreases, to devote more to enhancing the ocular contrast.

33. when the distance between the two eyes become larger,

    A. the correlation between left eye input S_L and right eye input S_R decreases,

    B. the ocular summation signal S_+ decreases

    C. the ocular difference signal S_- increases

    D. the ocular dominance columns in the brain should be weaker.

34. what are the intuition from the ocular efficient coding of two input signals, left and right eye inputs, to efficient coding in space, time, and color?

    A. Treat each space/time/color case as having two input nodes, at two spatial locations, two temporal instance, and from two cone types, then make an analogy to the two inputs from left and right eyes.

    B. g_+ versus g_- gain to the ocular summation versus ocular difference signals can now be extended intuitively as the gain to the signals summed over space, time, and cone type versus the gain to the contrast signal across space, time, and cone types.

    C. How g_+ and g_- change with input signal correlation and signal to noise can now be intuitively extended to how they should change with input signal to noise and input correlations in space, time, and color domains.

35. What is a principal component for spatial input signal $\mathbf{S}$ represented as a vector whose components are the image pixel values on evenly spaced pixels?

    A. It is a Fourier component of the spatial signal $\mathbf{S}$, one component for each Fourier frequencies k.

    B. it is a spatial average of all the image pixel values.

36. Let the i^{th} principal component of input $\mathbf{S}$ be the Fourier component with the corresponding spatial frequency k_i, please tick the correct statement below:

    A. the signal power, or variance, of this component is the corresponding, i^{th}, eigenvalue of the correlation matrix for $\mathbf{S}$.

    B. In natural visual scenes, the signal power, or variance, of this component decays with the magnitude of the Fourier frequency k_i, so that it is inversely related to the square of this frequency value.

37. Each principal component is a Fourier component of the original spatial signal $\mathbf{S}$. Let this component has frequency k and it has a gain $g(k)$ applied to it by the efficient coding recipe. Then

    A. because the power spectrum of the natural scene decays with the magnitidue $|k|$ of frequency k, the gain $g(k)$ increases with $|k|$ for small $|k|$, and then decreases with $|k|$ for larger $|k|$,

    B. the frequency magnitude $|k|$, where $g(k)$ peaks with $|k|$, is where the signal power of this Fourier component is comparable to the input noise power.

    C. $g(k)$ is to achieve whitening across the whole frequency range.

    D. $g(k)$ is related to the contrast sensitivity function of the neurons and visual behavior.

38. In efficient spatial coding of visual spatial signal $\mathbf{S}$, the multiplexing matrix U in the encoding matrix $\mathsf{K} = \mathsf{K}_o\mathsf{g}\mathsf{U}$ is

A. the inverse Fourier transform matrix, i.e., $\mathsf{U} = (\mathsf{K}_o)^{-1}$.

B. an identity matrix.

39. combining the three conceptual steps (principal component decomposition by Fourier transform K_o, gain control g, multiplexing U by inverse Fourier transform) for the efficient spatial coding $\mathsf{K} = \mathsf{K}_o\mathsf{g}\mathsf{U}$ on the retina does

A. gives the output $\mathbf{O}$, represented as an image, is a spatially filtered version of the input spatial image $\mathbf{S}$.

B. The spatial filter from the $\mathbf{S}$ image to the $\mathbf{O}$ image is a band-pass filter described by the filter $g(k)$ for the gain control step.

40. After the efficient coding of the spatial signal $\mathbf{S}$ to encoded response $\mathbf{O}$, we can obtain the correlation between different components of the $\mathbf{O}$ vector. Let $\mathbf{O}$ be represented as an image after filtering the input image $\mathbf{S}$ by the efficient spatial receptive field $K(x)$, then,

A. $\mathbf{O}$ between any two different pixels are completely uncorrelated with each other.

B. $\mathbf{O}$ at neighboring pixels are correlated with each other, if they are sufficient close, e.g., the distance between the pixels are smaller than size of the spatial filter $K(x)$.

C. $\mathbf{O}$ at different pixels are approximately not correlated with each other, if they are sufficiently far from each other.

41. When the signal to noise in the input $\mathbf{S}$ decreases, the size of the efficient coding receptive field $K(x)$

A. increases.

B. decreases.

C. stays unchanged.

42. Efficient temporal coding is analogous to efficient spatial coding except for some differences. These difference include:

A. space stretches in two dimensions, time extends in one dimension only.

B. need to ensure temporal causality in the encoding.

C. translation invariance in the signal correlation structure in space does not hold in time.

D. in the power spectrum, the signal power does not decay with temporal frequency.

43. In temporal coding, what is the analogous of the size of the spatial receptive field?

A. temporal extent of the impulse response function, which is the temporal receptive field filter

B. amplitude of the temporal receptive field filter

44. When the input signal to noise decreases, how should the efficient temporal filter change?

A. its temporal integration duration should increase

B. it should increase its amplitude to extract more input signals.

C. it should put less emphasis on extracting the difference between input signals at different times.

45. At what signal to noise condition is predictive temporal encoding more applicable?

A. high signal to noise

B. low signal to noise.

46. considering a trichromatic situation to encode input color, ignore spatial, temporal, and ocular input feature dimensions, what is the dimension of the input signal $\mathbf{S}$?

A. 3

B. N, as many as the input pixels.

47. Which of the following is more analogous to ocular coding (to encode inputs **S** from the left and right eyes)?

    A. dichromatic color vision

    B. trichromatic color vision

48. What are the principal components for 3-d color input **S** in trichormatic color vision?

    A. luminance, red-green chromatic, and blue-yellow chromatic

    B. luminance and chromatic channels.

49. When the noise is negligible, the gain to the chromatic channel should be

    A. lower than the gain to the luminance channel

    B. higher than the gain to the luminance channel.

50. Let us say that we have a retinal ganglion cell that has its receptive field property already fixed by the statistics of natural scenes, such that it no longer adapts its receptive field properties to the changes in input statistics. Let us measure this neuron's receptive field properties. We can measure it using luminance gray scale images, such that at each location red signal is always the same as the green signal input. In other words, red-green =0 across image locations, but red+green can vary across image locations to give us a luminance image. We can also measure this neuron's receptive field using chromatic images with no luminance contrast, such that across the spatial locations in the visual input, red+green signal is constant, but red-green signal varies to give a chromatic image. The receptive field measured using luminance images is larger or smaller than the receptive field measured using chromatic images? Please use efficient coding knowledge to deduce the answer.

    A. larger

    B. smaller

    C. the same size.

51. In retinal ganglion cells tuned to color, we have, for example, a neuron excited by red color in the center of the receptive field but inhibited by green color in the surround part of the receptive field. the radius for this center part and the radius for the surround part are likely to be approximately

    A. the radius of the center part of the receptive field meausred using the luminance signal, and the radius of the center part of the receptive field measured using the chromatic signal.

    B. the radius of the center and surround parts of the receptive field for the luminance signal.

    C. the radius of the center and surround parts of the receptive field for the chromatic signal.

52. V1 neurons are tuned to spatial frequencies, different neurons are tuned to different spatial frequency bands. The frequency bandwidth for neurons tuned to higher frequencies is wider, when measured in terms of the frequency difference between the high end and low end of the frequency band. In particular, this bandwidth is proportional to the mean frequency of the frequency band. What does this do?

    A. this helps to achieve scale invariance.

    B. This makes the shapes of the receptive fields similar across different scales (for V1 neurons' receptive fields). (The neurons can differ in the size and sensitivities of their receptive fields.)

53. In primates, the number of neurons in V1 is

    A. 10 times as many as the number of retinal ganglion cells

    B. 100 times as many as the number of retinal ganglion cells

    C. the same as the number of retinal ganglion cells.

54. What is a likely reason for the overcompleteness to be problematic for efficient coding?

A. overcompleteness implies that V1 has a very redundant representation.

B. more neurons incur more cost.

55. Which one of the following is a parameter needed to describe a gabor filter?

A. amplitude A of the filter.

B. size sigma of the gaussian envelop,

C. frequency and phase of the sinusoidal part of the gabor filter.

56. Because visual signals are correlated in natural scenes, summing signals across neighboring spatial location, across nearby time instances, across different cone types, and/or across two different eyes, tend to give statistically higher signal power than contrasting signals (or taking a difference between signals) at different coordinate values in space, time, color, or eye of origin. Meanwhile, contrasting or differentiating signals is the basis for feature selectivity (selectivity to space, time, frequency, orientation, color, disparity, etc).

A. according to efficient coding, this is why gains to visual input tend to be weaker when the signals are differentiated in too many feature dimensions (to achieve multiple-dimensional feature selectivity).

B. This is why V1 has few cells tuned to more than two feature dimensions.

C. This cannot explain why humans cannot see flickering color checkerboards when the flickering frequency is too high.