Question 1

Not yet answered

Marked out of 1.00

Flag question

Which of these statements about Dropout is FALSE:

模拟3一个网络构架集合

Select one:

- a. Dropout simulates an ensemble of network architectures
- b. Dropout helps to prevent overfitting
- c. Dropout encourages redundancy 冗余
- d. Dropout encourages the weight values to be small ✓

Question 2

Not yet answered

Marked out of 1.00

Flag question

Whit type of Autoencoder explicitly forces the hidden features not to change much when the inputs are slightly altered?

Select one:

- a. Variational Autoencoder
- b. Sparse Autoencoder
- c. Denoising Autoencoder
- d. Contractive Autoencoder ✓

Question 3

Not yet answered

Marked out of 1.00

Flag question

The best way to deal with the problem of temporal correlations in Deep Q-Learning is:

Select one:

- a. Cross Entropy Minimization
- b. Increased Momentum
- c. Experience Replay ✓
- d. Back Propagation Through Time

Question 4

Not yet answered

Marked out of 1.00

Flag question

Considering a Singular Value Decomposition $X = U\,S\,V^T$, what are the special properties of matrices U, S and V ?

Select one:

- a. U is orthogonal, V is upper triangular and S is symmetric.
- b. U, V are upper triangular, and S is diagonal.
- c. U, V are symmetric and S is orthogonal.
- d. U, V are unitary and S is diagonal. ✓

Question 5
Not yet answered
Marked out of 1.00
Flag question

Two common methods for unsupervised pre-training of neural networks are:

Select one:

- a. Deep Boltzmann Machine and Bayesian Inference
- b. Weight Initialization and Autoencoder
- c. Bayesian Inference and Weight Initialization
- d. Autoencoder and Deep Boltzmann Machine ✓

Question 6
Not yet answered
Marked out of 1.00
Flag question

When training on linearly separable data using the Perceptron Learning Rule, what will happen if both the learning rate and the initial weights are scaled up by a large factor?

Select one:

- a. The data will be learned successfully, but in a larger number of epochs
- b. The data will be learned successfully, in a smaller number of epochs
- c. The data will be learned successfully, in about the same number of epochs ✓
- d. Learning may become unstable and fail to converge

Question 7
Not yet answered
Marked out of 1.00
Flag question

Reinforcement Learning is when an agent is:

Select one:

- a. presented multiple times (over time) with the same examples of inputs and their target outputs
- b. only presented with the inputs and not target outputs, so it aims to find structure in these inputs
- c. not presented with target outputs, but instead given a reward signal that it aims to maximize ✓
- d. presented once with examples of inputs and their target outputs

unsupervised learning.

Question 8
Not yet answered
Marked out of 1.00
Flag question

When using Batch Normalization, in the Testing phase, the Mean and Variance of the activations at each node are typically:

Select one:

- a. pre-computed from the training set
- b. estimated using running averages
- c. either of the above ✓
- d. none of the above

Question 9
Not yet answered
Marked out of 1.00
Flag question

When comparing a Hopfield Network with a Boltzmann Machine, which statement is FALSE?

Select one:

- a. The range of activations is {-1,1} for one model and {0,1} for the other
- b. One model is used for retrieval, the other for generation
- c. The formula for the energy function is different for the two models
- d. The updates are deterministic for one model, and stochastic for the other

Question 10
Not yet answered
Marked out of 1.00
Flag question

The Context Layer in a Simple Recurrent Network:

Select one:

- a. is computed from the current input and the previous hidden layer
- b. is comprised of the inputs in a sliding window around the current timestep
- c. is a copy of the hidden layer from the previous timestep
- d. is computed from the current input and the previous output

Question 11
Not yet answered
Marked out of 1.00
Flag question

Which statement about word2vec is FALSE?

Select one:

- a. Representations for the same word at the input and output layers are different
- b. It aims to maximise the log probability of a word, based on the surrounding words
- c. The tanh activation function is used at the hidden nodes
- d. Performance improves if frequent words are sampled less often

不使用tanh激活函数.

Question 12
Not yet answered
Marked out of 1.00
Flag question

Which of these is NOT a method for dealing with the problem of vanishing or exploding gradients?

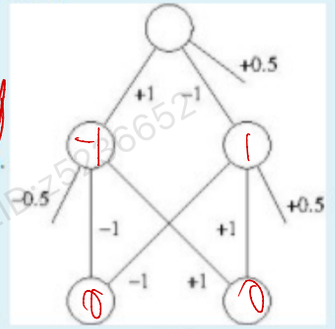
Select one:

- a. Batch Normalization
- b. Rectified Linear Unit
- c. Weight Initialization
- d. Conjugate Gradients

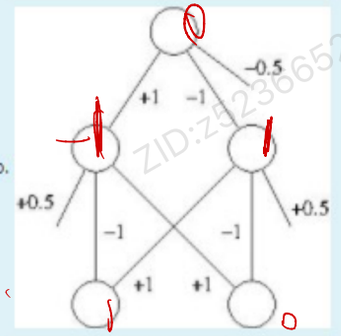
If 0=FALSE and 1=TRUE, which of these networks (with threshold activations at both the hidden and output layer) correctly computes the XOR function of two inputs?

Select one:

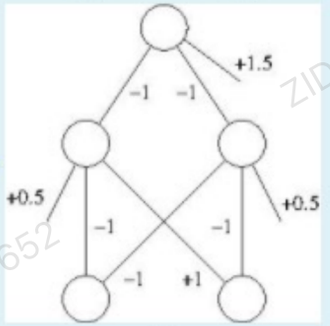
a.



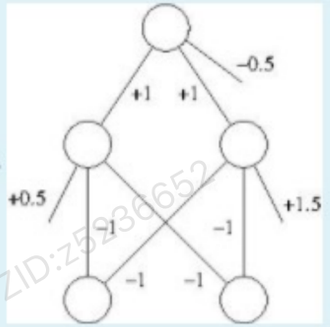
b.



c.



d.



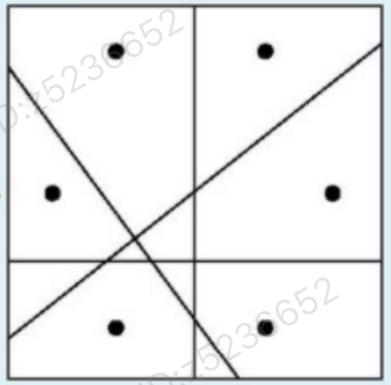
Consider a fully connected feedforward neural network with 6 inputs, 2 hidden units and 4 outputs, using tanh activation at the hidden units and sigmoid at the outputs. Suppose this network is trained on the following data, and that the training is successful.

Item	Inputs	Outputs
	123456	1234
1.	100000	0001
2.	010000	0011
3.	001000	0100
4.	000100	1010
5.	000010	1011
6.	000001	1110

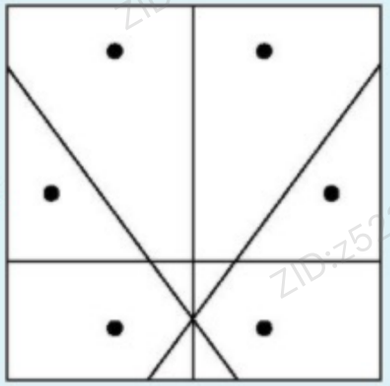
Which of these diagrams correctly shows a point in hidden unit space corresponding to each input, and, for each output, a line dividing the hidden unit space into regions for which the value of that output is greater/less than one half ?

Select one:

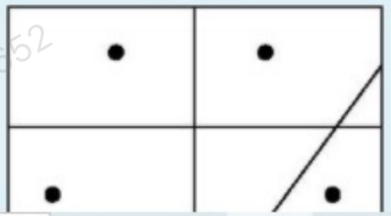
○ a.



○ b.



○ c.



Question **13**

Not yet answered

Marked out of 2.00

Flag question

Consider a Perceptron whose output is given by $h(w_0+w_1x_1+w_2x_2)$, where x_1, x_2 are inputs and $h()$ is the Heaviside (step) function.

Assume this Perceptron is being trained on the data in the following table, and that the current values of the weights are w_0= 0.5, w_1= 1 and w_2= -2.

Training Example	x_1	x_2	Class
(a)	−1	−1	Neg
(b)	2	1	Neg
(c)	−2	2	Pos

If the Perceptron Learning Rule is applied to the current weights, using training item (b) and a learning rate of η=1.0, the new values for w_0, w_1 and w_2 at the end of this training step will be:

* w_0: -0.5
* w_1: 1
* w_2: -3

0.5 + 2 × 1 − 2 × 3 = 0.5

−0.5 −1 −3

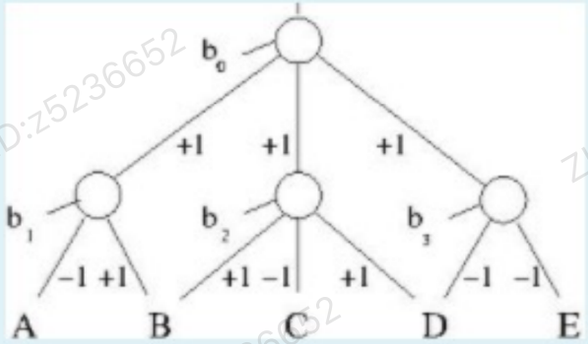
Question **14**

Not yet answered

Marked out of 2.00

Flag question

Consider the following multi-layer perceptron, using the threshold activation function, and assume that TRUE is represented by 1; FALSE by 0.



For which values of the biases b_0, b_1, b_2 and b_3 would this network compute the logical function

(¬A ∨ B) ∧ (B ∨ ¬C ∨ D) ∧ (¬D ∨ ¬E)

* b_0= -2.5
* b_1= 0.5
* b_2= 0.5
* b_3= 1.5

$P(P|B)=20\%$ $P(B)=\frac{3}{5}$ $P(\neg B)=\frac{2}{5}$

$P(B|P)$

$P(P)=P(P|B)\times P(B)+P(P|\neg B)\times P(\neg B)$

$=\frac{1}{5}\times\frac{3}{5}+\frac{7}{10}\times\frac{2}{5}=\frac{10}{25}=\frac{2}{5}$.

Question 15
Not yet answered
Marked out of 2.00
Flag question

Fred's Flower Shop buys 40% of its plants from Nursery A and 60% from Nursery B. Among the plants grown at Nursery A, 30% of them produce white flowers and 70% produce pink flowers. Among the plants grown at Nursery B, 80% of them produce white flowers and 20% produce pink flowers. If a plant in Fred's Flower Shop produces pink flowers, what is the probability that it was grown in Nursery B?

Answer: 0.3

$P(B|P)=\frac{P(P|B)\times P(B)}{P(P)}=\frac{1}{5}\times\frac{3}{5}\Big/\frac{2}{5}=\frac{3}{10}$

Question 16
Not yet answered
Marked out of 3.00
Flag question

Consider these two probability distributions on the same space Ω = {A, B, C, D, E}

p = (1/2, 1/4, 1/8, 1/16, 1/16)
q = (1/8, 1/16, 1/4, 1/2, 1/16)

$\frac{1}{2}+\frac{1}{4}+\frac{3}{8}+\frac{1}{16}+\frac{1}{16}=1.875$.

Compute (correct to at least two decimal places):

* The Entropy H(p): 1.875 $H(P)=-\frac{1}{2}(\log_2\frac{1}{2})-\frac{1}{4}(\log_2\frac{1}{4})-\frac{1}{8}(\log_2\frac{1}{8})-\frac{1}{16}(\log_2\frac{1}{16})-\frac{1}{16}(\log_2\frac{1}{16})$

* The KL-Divergence $D_{KL}(p \| q)$: 1.1875

$D_{KL}(P\|q)=\frac{1}{2}(\log_2\frac{1}{2}-\log_2\frac{1}{8})+\frac{1}{4}(\log_2\frac{1}{4}-\log_2\frac{1}{16})+\frac{1}{8}(\log_2\frac{1}{8}-\log_2\frac{1}{4})+$

$\frac{1}{16}(\log_2\frac{1}{16}-\log_2\frac{1}{32})+\frac{1}{16}(\log_2\frac{1}{16}-\log_2\frac{1}{16})$

$=1+\frac{1}{4}\times2+\frac{1}{8}\times(-1)+\frac{1}{16}\times(-3)$

$=1+0.5-\frac{1}{8}-\frac{3}{16}$

Question 18

Not yet answered

Marked out of 3.00

Flag question

Consider a neural network trained using softmax for a classification task with three classes 1, 2, 3. Suppose a particular input is presented, producing outputs

$z_1 = 1.3$, $z_2 = 2.4$, $z_3 = 3.7$

Assuming the correct class for this input is Class 2, and that Prob(2) is the softmax probability of the network choosing Class 2, compute the following, to two decimal places:

* d(log Prob(2))/dz1 = −0.07
* d(log Prob(2))/dz2 = 0.80
* d(log Prob(2))/dz3 = −0.73

$\log \text{Prob}(2) = e^{2.4} - \log(e^{1.3} + e^{2.4} + e^{3.7})$

$d(\log \text{Prob}(2))/dz_1 = -\frac{e^{1.3}}{e^{1.3} + e^{2.4} + e^{3.7}}$

$d(\log \text{Prob}(2))/dz_2 = 1 - \frac{e^{2.4}}{e^{1.3} + e^{2.4} + e^{3.7}}$

$d(\log \text{Prob}(2))/dz_3 = -\frac{e^{3.7}}{e^{1.3} + e^{2.4} + e^{3.7}}$

Question 19

Not yet answered

Marked out of 3.00

Flag question

Consider a convolutional neural network which takes as input a 42-by-54 color image (i.e. with three channels R, G, B). The first convolutional layer has 16 filters that are 6-by-6, with stride 3 and no zero-padding.

Compute the number of:

* weights per neuron in this layer (including bias): 109 $1 + 6 \times 6 \times 3$
* neurons in this layer: 3536 $13 \times 17 \times 16 = 3536$.
* connections into the neurons in this layer: 385424 $13 \times 17 \times 16 \times 109$
* independent parameters in this layer: 1744

$(1 + (42 - 6)/3) \times (1 + (54 - 6)/3)$

13×17

Question 20

Not yet answered

Marked out of 3.00

Flag question

Consider a Hopfield Network with the following weight matrix W:

0 0 -1 0 0
0 0 0 0 -1
-1 0 0 +1 0
0 0 +1 0 +1
0 -1 0 +1 0

For each of the following vectors, state whether it is Stable or Not Stable for this network:

* [−1 +1 +1 −1 +1]: ~~Not~~ Stable
* [+1 +1 −1 −1 −1]: stable
* [+1 +1 −1 +1 +1]: Not stable

[−1, −1, 2, 2, 0]

[1, 1, −2, −2, −2]

[1, −1, 0, 0, 0]

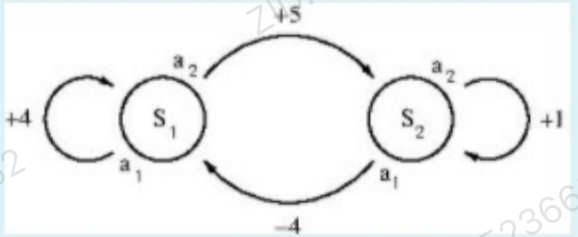
Question 21

Not yet answered

Marked out of 6.00

Flag question

Consider an environment with two states S = {S_1, S_2} and two actions A = {a_1, a_2}, where the (deterministic) transitions δ and reward R for each state and action are as follows:



Assuming a discount factor of γ = 0.5, determine:

* $\pi^*(S_1)$= a1
* $\pi^*(S_2)$= a2

Again assuming γ = 0.5, compute these values (correct to two decimal places):

* $Q^*(S_1,a_1)$ 8
* $Q^*(S_1,a_2)$ 6
* $Q^*(S_2,a_1)$ 0
* $Q^*(S_2,a_2)$ 2

If γ is allowed to vary between 0 and 1, for which range of values of γ is this policy optimal (correct to two decimal places)?

* Minimum value of γ: 0.25
* Maximum value of γ: 0.625

(1) a_1a_1:

$V(S_1) = 4 + 0.5 \times V(S_1)$ $\quad V(S_1) = 8$

$V(S_2) = -4 + 0.5 \times V(S_1)$ $\quad V(S_2) = 0$

a_1a_2:

$V(S_1) = 4 + 0.5 \times V(S_1)$ $\quad V(S_1) = 8$

$V(S_2) = 1 + 0.5 \times V(S_2)$ $\quad V(S_2) = 2$

a_2a_1

$V(S_1) = 5 + 0.5 \times V(S_2) =$

$V(S_2) = -1 + 0.5 \times V(S_1)$

$0.5V(S_1) = 2.5 + 0.25V(S_2)$

$0.5V(S_1) - 0.25V(S_2) = 2.5$

$V(S_1) = 4$

$V(S_2) = -2.$

Ans:

$V(S_1) = 5 + 0.5V(S_2)$

$V(S_2) = 1 + 0.5V(S_2)$

$V(S_2) = 2, \quad V(S_1) = 6$

$V(S_2) - 0.5V(S_1) = -4$

$V(S_2) - 0.25V(S_2) = -1.5$

$0.75V(S_2) = -1.5$

$V(S_2) = -2$

$V(S_1) = 4$

(c) $Q(S_1, a_2) = 5 + 0.5 \times V(S_2) = 5 + 0.5 \times 2 = 6$

$Q(S_2, a_2) = -4 + 0.5 \times V(S_1) = -4 + 0.5 \times 8 = 0$

(d) (a) (a)

$V(S_1) = 4 + \gamma V(S_1)$

$V(S_1) = \frac{4}{1-\gamma}$

$V(S_2) = -4 + rV(S_1)$

$V(S_2) = \frac{4r}{1-r} - \frac{4-4r}{1-r} = \frac{8r-4}{1-r}$

(a_2, a_1)

$V(S_1) = 5 + rV(S_2)$

$V(S_2) = -4 + rV(S_1)$

$V(S_1) = \frac{4r - 5r^2 + 5r^2 - 5}{r^2 - 1}$

$= \frac{4r - 5}{r^2 - 1}$.

(a_1, a_2)

$V(S_1) = 4 + rV(S_1)$

$V(S_2) = 1 + rV(S_2)$

$V(S_2) = \frac{4 - 4r}{r^2 - 1}$.

$V(S_1) = \frac{4}{1-r}$

$V(S_2) = \frac{1}{1-r}$

(a_2, a_2)

$V(S_1) = 5 + rV(S_2)$

$V(S_2) = 1 + rV(S_2)$

$V(S_1) = \frac{5-4r}{1-r}$

$V(S_2) = \frac{1}{1-r}$

S_1:

$$\frac{4}{1-r} > \frac{5-4r}{1-r}$$

$r > \frac{1}{4}$

$$\frac{4}{1-r} > \frac{4r-5}{r^2-1}$$

$$\frac{4}{1-r} > \frac{4r-5}{(r+1)(r-1)}$$

$-4r-1 > 4r-5$

$$\frac{-4}{r-1} > \frac{4r-5}{(r+1)(r-1)}$$

$-8r > -4$

$r < \frac{1}{2}$

即 $\frac{1}{4} < r < \frac{1}{2}$.

S2:

$$\frac{1}{1-r} > \frac{8r-4}{1-r}$$

$$1 > 8r-4$$

$$5 > 8r \qquad r < \frac{5}{8}$$

$$\frac{1}{1-r} > \frac{4-5r}{r^2-1}$$

$$\frac{1}{1-r} > \frac{4-5r}{(r-1)(r+1)}$$

$$-\frac{1}{r-1} > \frac{4-5r}{(r-1)(r+1)}$$

$$-r-1 > 4-5r$$

$$4r > 5$$

$r \to \frac{\pi}{4}$

因为 $0<r\leq 1$ 所以由上可知 $\frac{\pi}{4}<t<\frac{5}{8}$.