General Principles of Human and Machine Learning: Pop Quiz #1

Name: **Date:**   **Grade:**

—---

1. What are Thorndike's two laws? Name and describe each, giving real-world examples of the laws in action.

2. What is the key difference between classical (Pavlovian) conditioning and operant (instrumental) conditioning?

3. How are the weights in the Rescorla-Wagner model updated? You don't need to write out the equation, but describe the general principle

4. When multiple cues predict reward, they compete for credit. Overshadowing and blocking are examples of this phenomenon. Complete the following descriptions
   a. Overshadowing: An animal is trained to associate both sound and light cues with reward. When presented with each cue individually, the response is _____ than when both cues are presented together
   b. Blocking: If an animal is first trained to associate a light cue with reward, and then later both light and sound cues with reward, which cue will have a stronger association with reward?

5. Draw a McCulloch & Pitts neuron with labeled weights and threshold that solve the following problems below. You can use the table on the next page. For convenience, the firing function is provided

$$f(\mathbf{x}) = \begin{cases} 1 & \text{if} \sum w_i x_i \geq \theta \\ 0 & \text{else} \end{cases}$$

   a. AND: all inputs need to be activated for it to fire
   b. OR: Neuron fires for any input is on
   c. NAND: Neuron fires when x1 is on AND x2 not on

AND	OR	NAND

6. Name one strength and one limitation of the Rosenblatt Perceptron learning rule

7. What limitations of the perceptron did Minsky & Parpert criticize? What generalization did they make?

8. What does a Physical Symbol System consist of and how do they interact to solve problems? [Hint: there are two main ingredients]

9. Name one strength and one weakness for:
   a. Symbolic AI:

   b. Subsymbolic AI
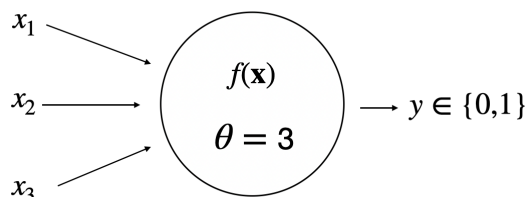
10. Explain the difference between S-S and S-R learning

Pop Quiz #1: Answer Key (DO NOT DISTRIBUTE)

1. i) The law of effect states that actions that successful actions performed in the past will be more likely to be repeated in the future. Examples include
   a. Getting encouragement for good grades (and then studying harder)
   b. Enjoying spending time with friend X, and then hanging out with them more often
   c. Having a good meal and a restaurant, and then going there more often

   ii) The law of exercise states that actions performed in the past (regardless of reward or reinforcement) are more likely to be repeated in the future. Examples include
   a. Morning routine
   b. Commute to work
   c. Study/exercise routine

2. Classical conditioning describes the *passive* association of rewards/punishment with stimuli, whereas operant conditioning describes *active* selection of actions that are associated with rewards

3. Rescorla-Wagner weights are updated as a function of prediction error, based on the difference between the predicted and observed outcomes. (This is known as the delta-rule).

4. a) weaker. b) light

5. Examples below
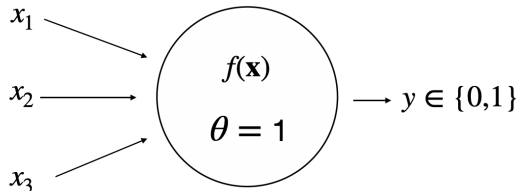
   AND function

   a.

b.

OR function

x_1

x_2 → $f(\mathbf{x})$ $\theta = 1$ → $y \in \{0,1\}$

x_3

NAND

x_1 $w_1 = 1$

$w_2 = -1$ $f(\mathbf{x})$ $\theta = 1$ → $y \in \{0,1\}$

x_2

$w_i \in \{-1,1\}$
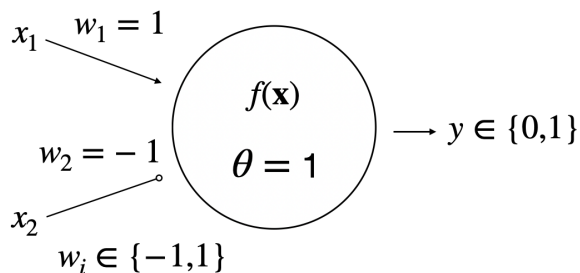
c.

6. Strength: guaranteed to converge for linearly separable problems. Weakness: Even a single mislabeled training data prevents convergence

7. Single-layer perceptrons cannot solve non-linearly separable problems, such as XOR. The generalization they made is that larger networks with more layers would also have similar problems

8. A physical symbol system consists of **symbols** and **relations/rules.** Symbols can represent anything in the world, e.g., (Bagels), (ChatGPT), (Charley), …. While relations can be *predicates/adjectives* that describe qualities of a symbol, e.g., toasted(Bagel) or *verbs* describing how symbols interact with other symbols e.g., eat(Charley, Bagel)

9. Strengths and weaknesses (use your best judgment for answers)
   a. Symbolic:
      i. Strengths: interpretability due to explicit decision logic, rapid generalization, fast solutions generated in real time, etc…

   ii. Weaknesses: difficult to learn rules/symbols, limited generalization when rules don't apply exactly, scaling problem due to combinatorial explosion

  b. Subsymbolic:

   i. Strengths: Powerful learning and scalability, Universal approximation theorem (can learn any function under weak assumptions), automatically discover patterns that even experts may not be aware of, etc…

   ii. Weaknesses: limited interpretability due to distributed representation and no explicit decision logic, needs large amounts of data to generalize well, limited compositionality, etc…

10. S-R learning is the strengthening/weakening of stimulus-response relationships, whereas S-S learning describes the latent relational structure between stimuli, facilitating the construction of a map-like representation of the environment