1. Winnow has a better mistake bound than Perceptron.
   **Options:**
   - The function space is sparse.
   - The instance space is sparse.
   **Ans:** The function space is sparse.

2. We generate training examples with respect to each word in the text. We only consider one feature type,
   
   \[ w(-1) = w \]
   
   which will result in features that are ON when the word before the target word is ‘w’.
   The data set we have consists of a single sentence: “The girl is playing the piano and the boy is playing basketball.” (13 tokens, including the period). We use a special word “SSS” to represent the word before the first word. What is the dimensionality of the feature space when we only use the single feature type w(-1)?
   **Options:**
   - 9
   - 13
   - 12
   - 11
   - 8
   - 10
   **Ans:** 9

3. We generate training examples with respect to each word in the text. We only consider one feature type,
   
   \[ w(-1)\&w(+1) = w1, w2 \]
   
   which will result in features that are ON when the word before the target word is w1 and the word after the target word is “w2”. The data set we have consists of a single sentence: “The girl is playing the piano and the boy is playing basketball.” (13 tokens, including the period). We use a special word ”SSS” and “EEE” to represent the word before the first word and the word after the last word. What is the dimensionality of the feature space when we only use the single feature type w(-1)\&w(+1)?
   **Options:**
   - 12
4. The averaged perceptron is likely to generalize better than the basic perceptron when evaluated on new, previously unseen data.

**Options:**

- True
- False

**Ans:** True

5. When we compare a “standard” SGD with the AdaGrad augmentation of it, which of the following descriptions are correct for AdaGrad?

**Options:**

(a) It performs larger updates for infrequent features and smaller updates for frequent ones.

(b) It eliminates the need to tune the learning rate. Each feature has its own learning rate that is adapted by the data.

(c) The learning rate for a given feature is monotonically decreasing.

(d) The learning rate is fixed, unless the gradient is too large, in which case we make a small correction.

**Ans:** (a), (b) and (c)