CS449/549 Computational Learning Practice Problems 1 (Fall 2023)

True/False Circle one. Provide short explanations to your answers.

1. (T/F) The number of mistakes of the Perceptron algorithm does not depend on the sample size.
   True: the bound $1/\gamma^2$ *is independent of sample size.*

2. (T/F) Any sample in $\mathbb{R}^d$ of 1 point is linearly separable.
   True: if the labeled point is $(\vec{x}, y)$*, take* $\vec{w} = y\vec{x}$*.*

3. (T/F) For neural network, the weights can be initialized to random values.
   True: default safe choice in gradient descent optimization.

4. (T/F) A one-node neural network using $\tanh(z)$ activation can be used to learn linearly separable sample.
   True: $\tanh(\langle \vec{w}, \vec{x} \rangle)$ *approximates* $\mathtt{sign}(\langle \vec{w}, \vec{x} \rangle)$.

5. (T/F) Weighted Majority makes more mistakes than the best expert.
   True: by definition of "best" expert.

Perceptron Consider the following implementation of the Perceptron algorithm.

```
/* x[i] = i-th feature vector, y[i] = (-1/+1) label of x[i] */
w = zero_vector(dim)
Oops = True /* mistake alert */
while Oops:
  Oops = False
  for i = 1 to num_points:
    if y[i] != dot_product(w,x[i]):
      w = w + y[i]*x[i] /* overloaded vector addition and scalar multiply */
      Oops = True
return w
```

(T/F) The above algorithm is correct.
False: missing call to `sign` *in* `if` *statement.*