1. The problems that can arise from setting a learning rate t_k too small are that it could take a very long time to converge, or it might find a local minimum instead of a global minimum. If the learning rate is too big, we might jump over the global minimum. We could also end up oscillating around the minimum that is found.

2. (a) $P_A(z) > P_B(z)$
   There are two different inputs with the same weights, and two negatives make a positive when multiplied/divided together.

   (b) $P_B(u) > P_A(z)$
   The magnitude of u is bigger than z and both outputs end up being positive, and the only difference in the weights of A and B is the sign.

   (c) $P_C(v) = P_B(v)$
   The input data is (0, 0) so the predictions/results are the same

   (d) $P_D(z) > P_B(u)$
   D has a nonzero bias of 5 when applied makes $P_D(z) > P_B(u)$ while $u > z$ in magnitude originally

3. (a)

$$\begin{aligned} \text{Winter} &= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^T \\ \text{Spring} &= \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix}^T \\ \text{Summer} &= \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix}^T \\ \text{Fall} &= \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T \end{aligned}$$

   (b) The one-hot representation might be easier for the linear model because it assigns discrete values to classes. Unlike the other representations in the one-hot representations the dot product between two seasons will always be zero as the vectors are orthogonal to each-other. This way as the model makes predictions we can tell how much each class claims the data point.

   The drawbacks of the integer representation are that each one is best measured from distance to a single hyper-plane, meaning the representation needs to have a linear or ordered distribution. If a hyperplane bifurcates a class decision for a series with more than 2 classes meaning we could end up not knowing the precise probability of class it is claimed by. This may not work as well if the classes are not a series. The binary representation offers no "in-between" as a data-point can be described by only one absolute class. The one-hot representation offers the best things of both the previously mentioned representations without the same drawbacks.