1. *Show that tabular methods such as presented in Part I of this book are a special case of linear function approximation. What would the feature vectors be?*

   The feature would be an indicator of the state: the feature vector's length would be equal to the number of states, and $x(s_i)$ would be a vector that contains 1 at the i^{th} coordinate and 0 at the others. The weight vector would also have the same length as the number of states. The $i^t h$ weight would be our value estimate for s_i. When we're doing the update, the gradient of $x(s_i)$ is 0 everywhere except for coordinate i, where it's 1, so we would only update the weight for the current state.

2. *Why does (9.17) define* $(n+1)^k$ *distinct features for dimension* k*?*

   For each state s_j, we choose one of $n+1$ options: $s_j^0, s_j^1 \ldots, s_j^n$. There are k states, we make a choice for each them: we can do that $((n+1)^k)$-many ways.