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^th 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: $s_j^0, s_j^1 \dots, s_j^n$. There are k states, we make a choice for each them: we can do that $((n+1)^k)$-many ways.

3. *What n and $c_{i,j}$ produce the feature vectors*

$$x(s) = (1, s_1, s_2, s_1s_2, s_1^2, s_2^2, s_1s_2^2, s_1^2s_2, s_1^2s_2^2)^T?$$

$$\begin{aligned}
n &= 2 \\
c_{1,1} &= c_{1,2} = 0 \\
c_{2,1} &= 1, c_{2,2} = 0 \\
c_{3,1} &= 0, c_{3,2} = 1 \\
c_{4,1} &= 1, c_{4,2} = 1 \\
c_{5,1} &= 2, c_{5,2} = 0 \\
c_{6,1} &= 0, c_{6,2} = 2 \\
c_{7,1} &= 1, c_{7,2} = 2 \\
c_{8,1} &= 2, c_{8,2} = 1 \\
c_{9,1} &= 2, c_{9,2} = 2
\end{aligned}$$

4. *Suppose we believe that one of two state dimensions is more likely to have an effect on the value function than is the other, that generalization should be primarily across this dimension rather than along it. What kind of tilings could be used to take advantage of this prior knowledge?*

   Let's say it's the first dimension that is more likely to have an effect on the value function. In this case, the tiles should be short in the first dimension and long in the second dimension. This way, if we learn from an example (s_1, s_2), we update many (s_1, x) states and less (x, s_2) states.