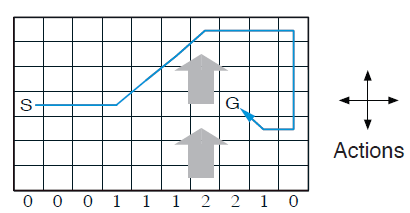
**Windy Gridworld with King’s Moves**

Shown inset below is a standard gridworld, with start and goal states, but with one difference: there is a crosswind running upward through the middle of the grid. The actions are King’s moves which include standard four—up, down, right, and left— and the diagonal moves. However, in the middle region, the resultant next states are shifted upward by a “wind,” the strength of which varies from column to column.

The strength of the wind is given below each column, in number of cells shifted upward. For example, if you are one cell to the right of the goal, then the action left takes you to the cell just above the goal. This is an undiscounted episodic task, with constant rewards of −1 until the goal state is reached.



**Optimal movement for the standard four movements.**

1. Solve this problem using SARSA. How much better can you do with the extra diagonal actions instead of the standard four? Can you do even better by including a ninth action that causes no movement at all other than that caused by the wind?
2. Solve this problem using Q-learning. How much better can you do with the extra diagonal actions instead of the standard four? Can you do even better by including a ninth action that causes no movement at all other than that caused by the wind?

*Note: Monte Carlo methods cannot easily be used here because termination is not guaranteed for all policies. If a policy was ever found that caused the agent to stay in the same state, then the next episode would never end. Online learning methods do not have this problem because they quickly learn during the episode that such policies are poor, and switch to something else.*

**Windy Gridworld with Stochastic Wind**

Re-solve the windy gridworld task with King’s moves, assuming that the effect of the wind, if there is any, is stochastic, sometimes varying by 1 from the mean values given for each column. That is, a third of the time you move exactly according to these values, as in the previous exercise, but also a third of the time you move one cell above that, and another third of the time you move one cell below that.

For example, if you are one cell to the right of the goal and you move left, then one-third of the time you move one cell above the goal, one-third of the time you move two cells above the goal, and one-third of the time you move to the goal.

1. Solve this problem using SARSA.
2. Solve this problem using Q-learning.

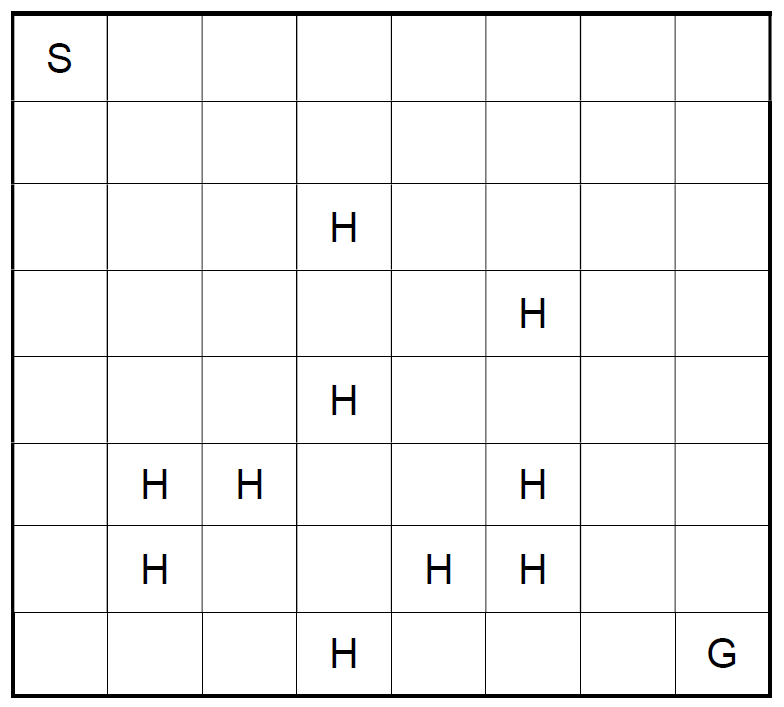
**Frozen Lake**

The game starts with the player at the starting point (**S**) of the frozen lake grid world and the goal (**G**) located at far extent of the world for the 8x8 environment below. The actions are the standard four—up, down, right, and left. However, the lake is slippery so the player may move perpendicular to the intended direction sometimes. The player will move in intended direction with probability of 1/3, and else will move in either perpendicular direction with equal probability of 1/3 in both directions. For example, if action is left, then:

* P(move left)=1/3
* P(move up)=1/3
* P(move down)=1/3

Movements that would take the player off the grid leave its location unchanged.

The player makes moves until they reach the goal or fall in a hole (**H**), both of which end the episode. This is an undiscounted episodic task with constant rewards of 0 except for when the goal is reached, which gives a reward of +1.



1. Solve this problem using SARSA.
2. Solve this problem using Q-learning.