Admin

- Midterm grades will be released today
  - Class did well
  - **no requesting changes to marks**
  - If you want to understand the answers, then come to my office hours
- Next Friday **Rich Sutton** will lecture about the role of experience in building AI

Mini Essay #2 Advice

- Mini-essay #2 topics are available here:
  - https://docs.google.com/document/d/1EOC5TQh-SKC5zYXrvvC44hR8wthwttrROKwyDPlUBz0
- Follow the instructions (if these things don’t make sense, then ask me about them):
  - Two paragraphs
  - Topic sentences
  - One idea per paragraph
- Common issue: marker could not understand what you were talking about
- Solution: follow the advice I gave in lecture, get a friend to read it over and ask them what they thought the essay was about

Review of Practice Quiz

Q&A about midterm answers

Exercise questions

Q0

Exercise 9.4 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?

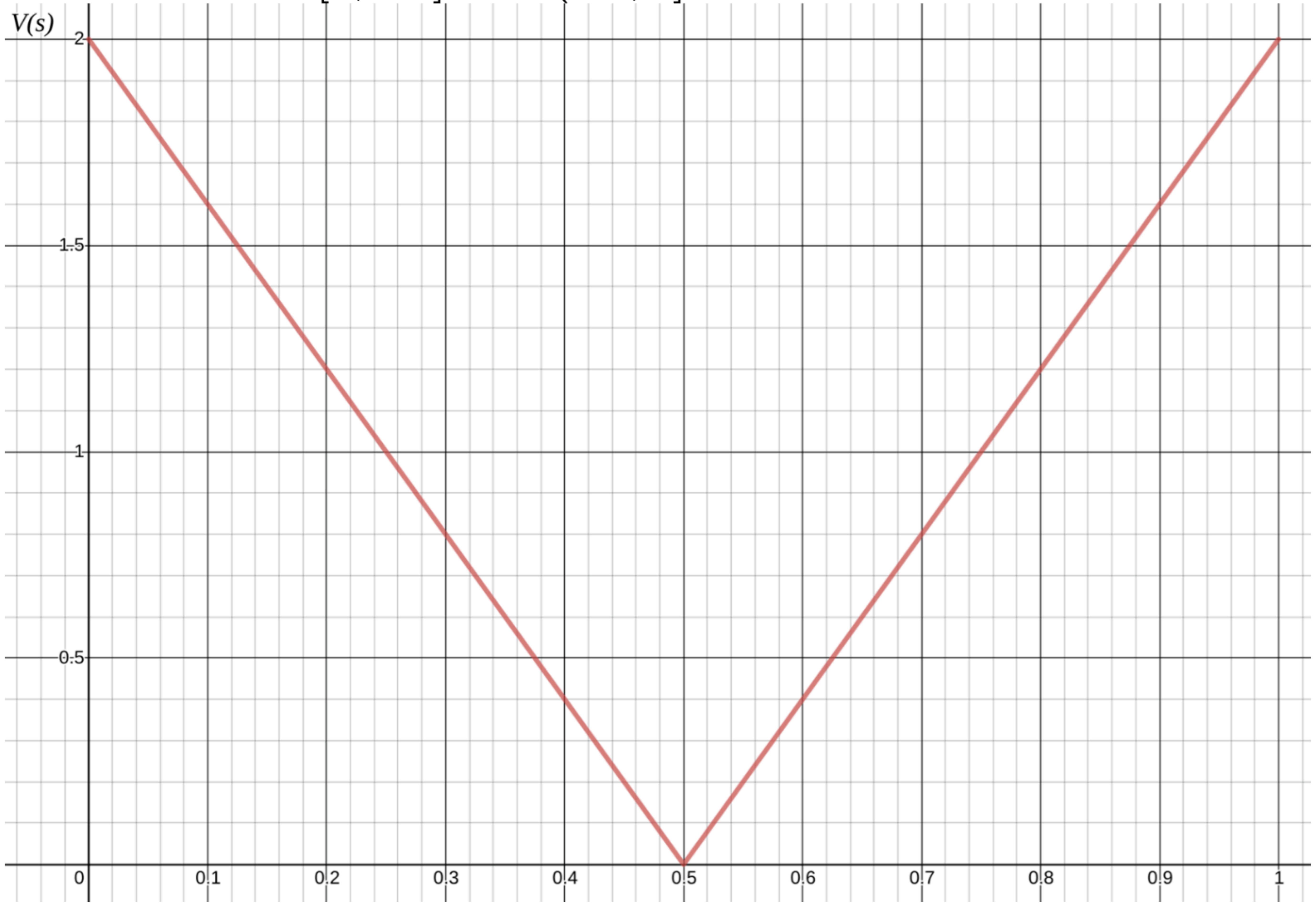
- Draw a picture with one tiling over a 2d state-space that could achieve this generalization?

Q1

3. Consider a problem with the state space, $\mathcal{S} = \{0, 0.01, 0.02, \cdots, 1\}$. Assume the true value function is

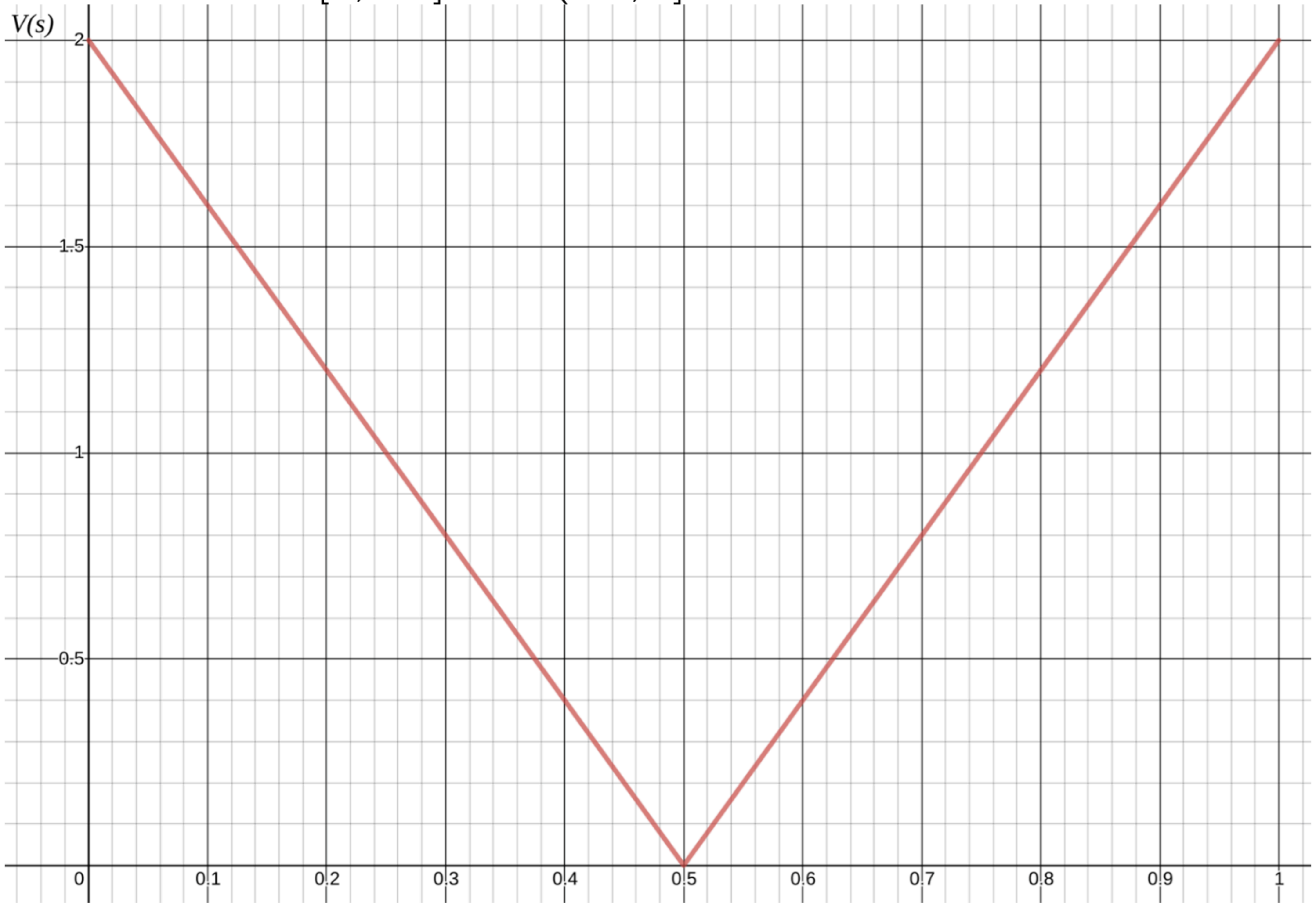
$$v_\pi(s) = 4|s - 0.5|$$

which is visualized below. We decide to create features with state aggregation, and choose to aggregate into two bins: $[0, 0.5]$ and $(0.5, 1]$.



(a) What are the possible feature vectors for this state aggregation?

to aggregate into two bins: $[0, 0.5]$ and $(0.5, 1]$.

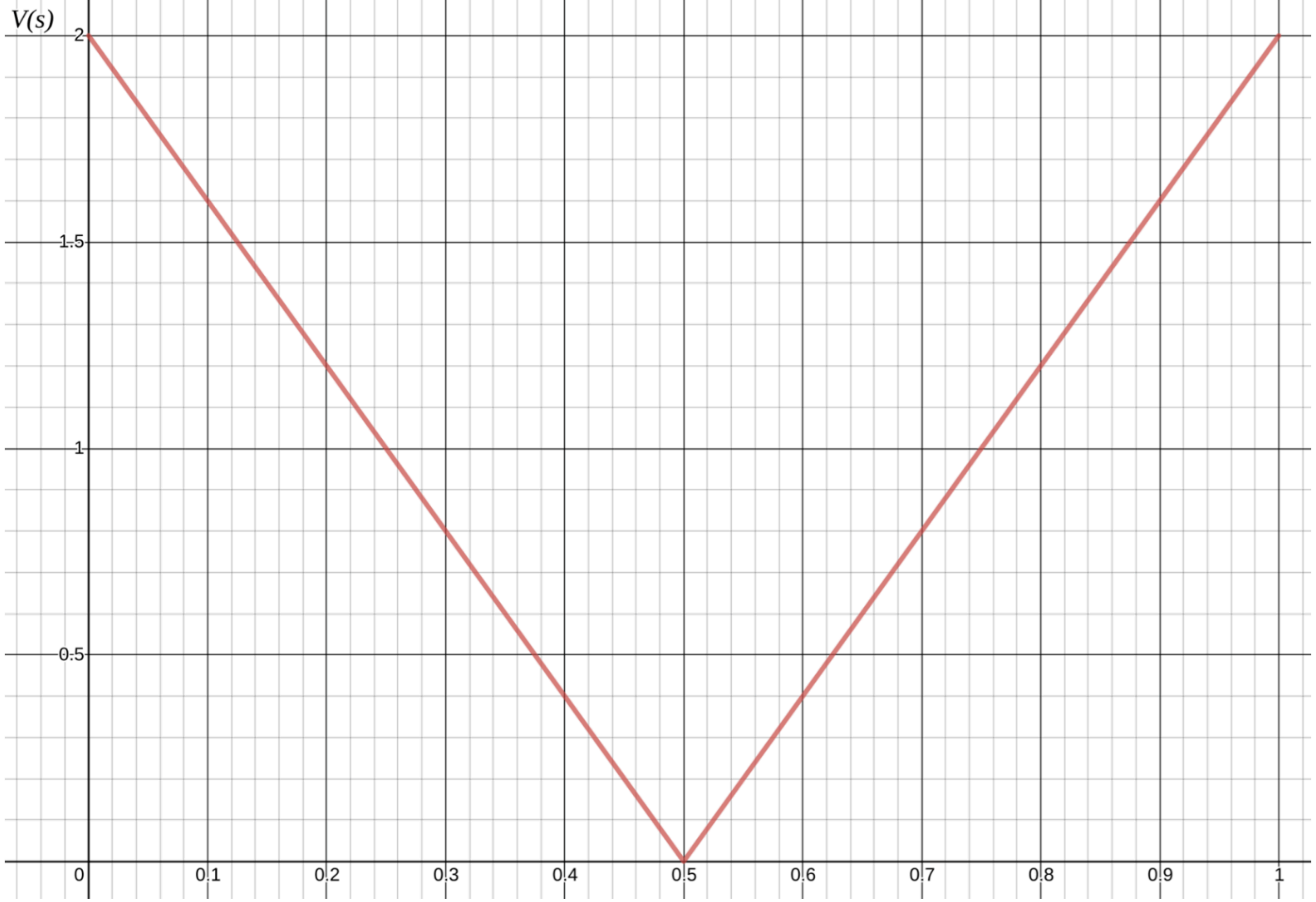


Q1

(a) What are the possible feature vectors for this state aggregation?

- Two bins, we can be in one or the other; we just “one-hot” encode those two possibilities

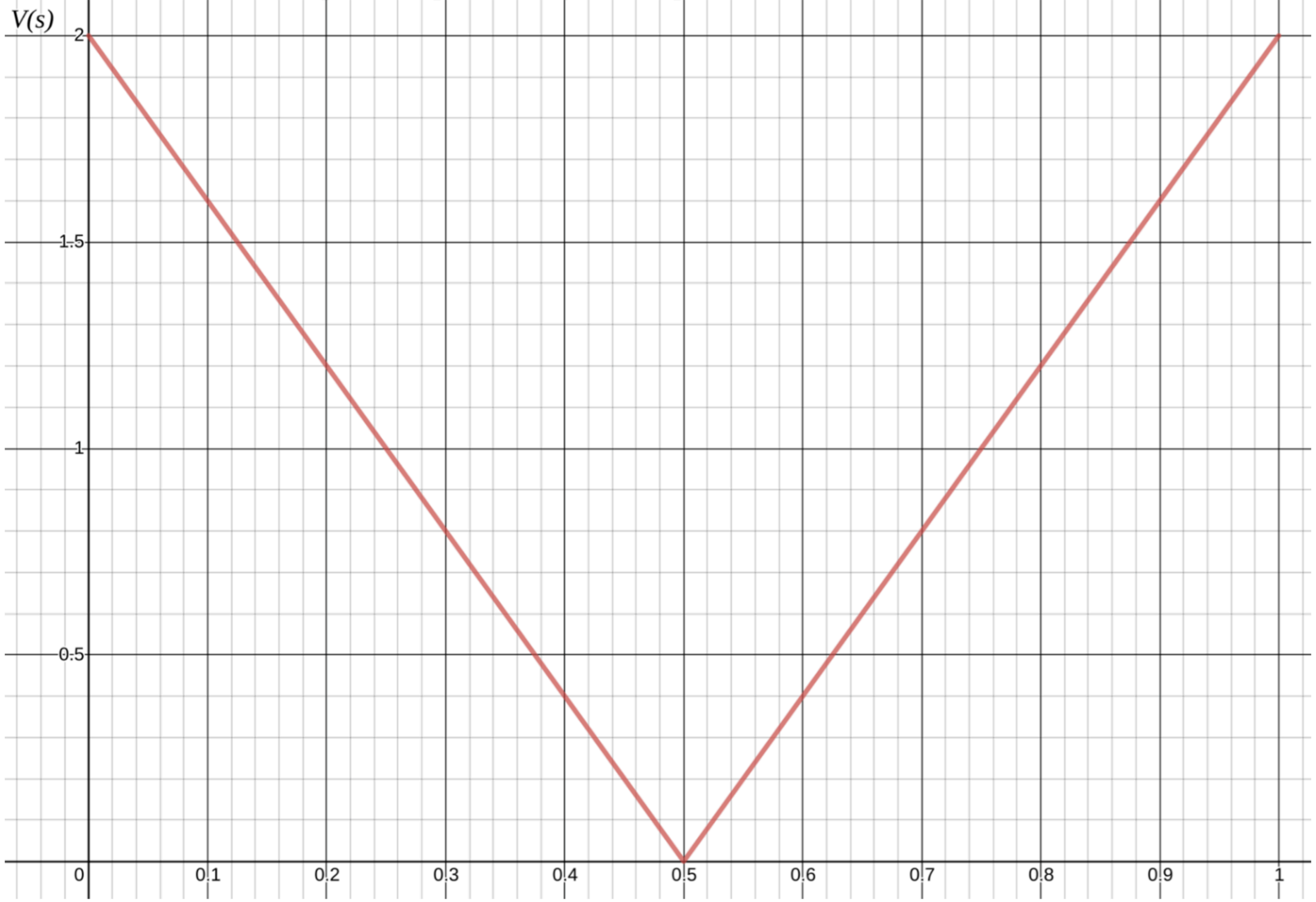
to aggregate into two bins: $[0, 0.5]$ and $(0.5, 1]$.



Q1

(b) Imagine you minimize the $\overline{\text{VE}}(\mathbf{w}) = \sum_{s \in \mathcal{S}} d(s)(v_\pi(s) - \hat{v}(s, \mathbf{w}))^2$ with a uniform weighting $d(s) = \frac{1}{101}$ for all $s \in \mathcal{S}$. What vector $\mathbf{w}$ is found?

to aggregate into two bins: $[0, 0.5]$ and $(0.5, 1]$.

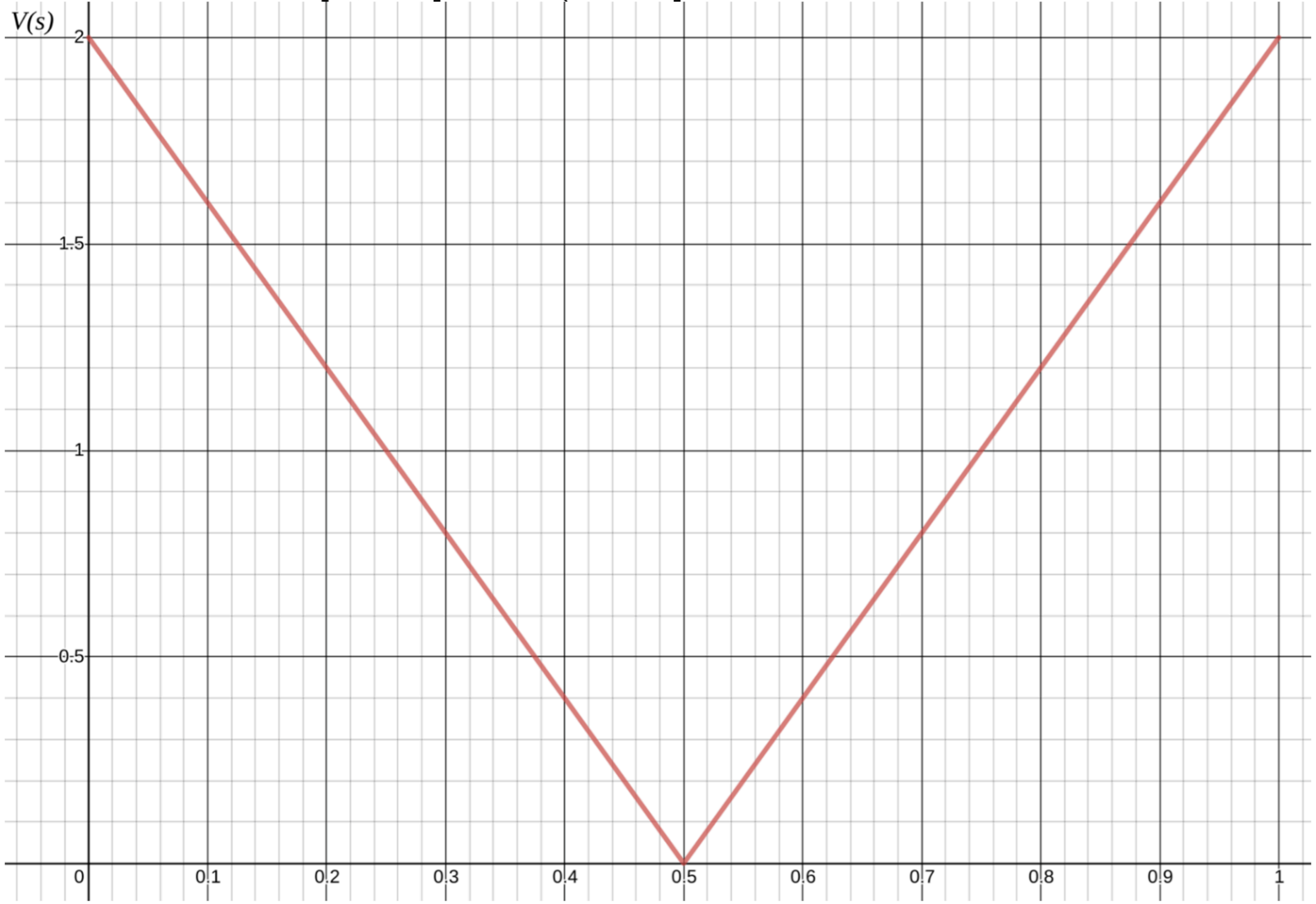


Q1

(b) Imagine you minimize the $\overline{\mathrm{VE}}(\mathbf{w}) = \sum_{s \in \mathcal{S}} d(s)(v_\pi(s) - \hat{v}(s, \mathbf{w}))^2$ with a uniform weighting $d(s) = \frac{1}{101}$ for all $s \in \mathcal{S}$. What vector $\mathbf{w}$ is found?

- If you you had one number to represent the function of the first side, ie fitting this line with a scalar?

to aggregate into two bins: $[0, 0.5]$ and $(0.5, 1]$.

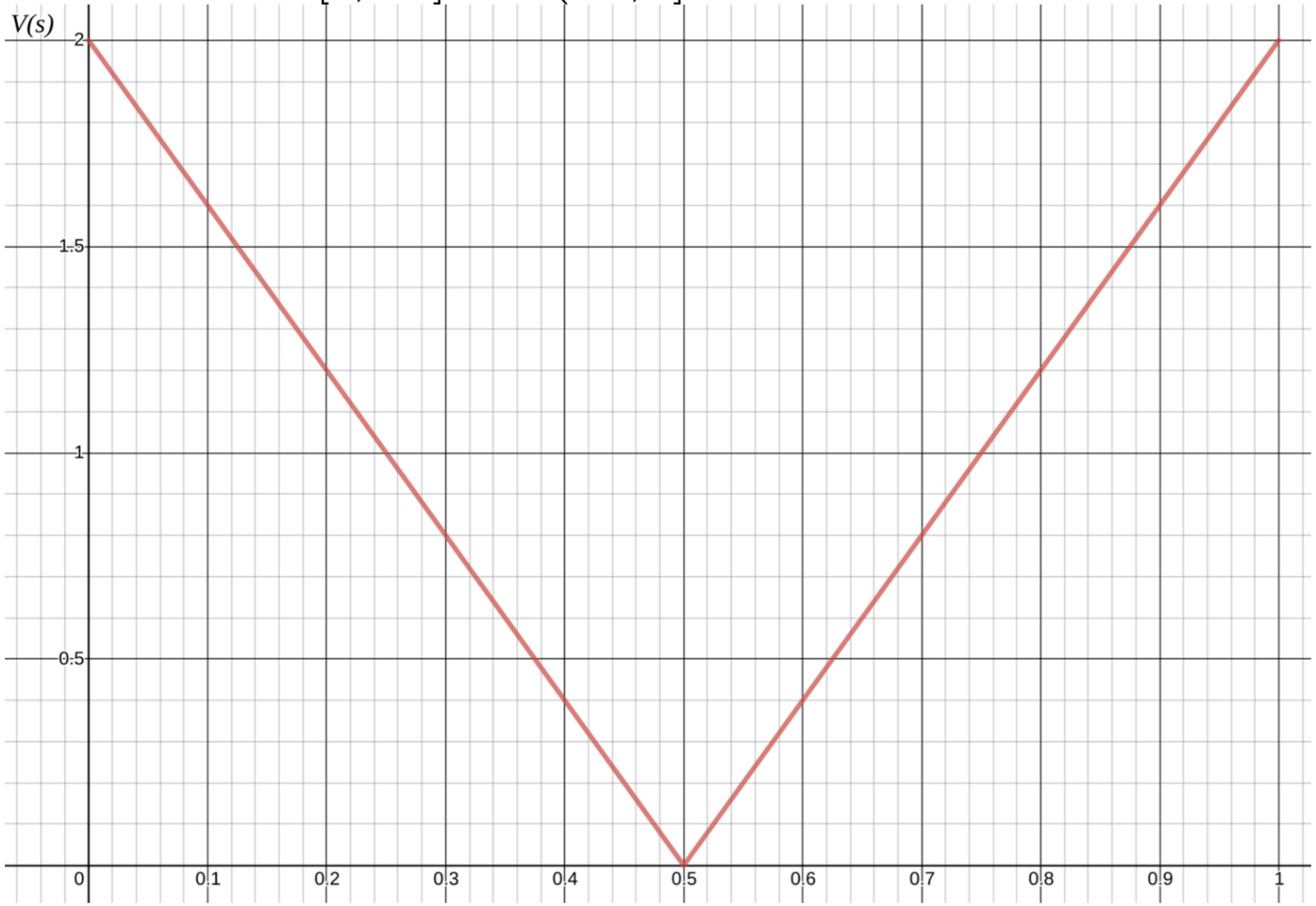


Q1

(c) Now, if the agent puts all of the weighting on the range $[0, 0.25]$, (i.e. $d(s) = 0$ for all $s \in (0.25, 1]$), then what vector $\mathbf{w}$ is found by minimizing $\overline{\mathrm{VE}}$?

to aggregate into two bins: $[0, 0.5]$ and $(0.5, 1]$.

Q1



(c) Now, if the agent puts all of the weighting on the range $[0, 0.25]$, (i.e. $d(s) = 0$ for all $s \in (0.25, 1]$), then what vector $\mathbf{w}$ is found by minimizing $\overline{\text{VE}}$?

- for s = 0 and s = 0.25: v(0) = 2 and v(0.25) = 1 …

Q2

4. Consider the following general SGD update rule with a general target U_t

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \alpha \left[U_t - \hat{v}(S_t, \mathbf{w}_t)\right] \nabla \hat{v}(S_t, \mathbf{w_t}).$$

   Assume we are using linear function approximation, *i.e.* $\hat{v}(S, \mathbf{w}) = x(S)^\top \mathbf{w}$.

   (a) What happens to the update if we scale the features by a constant and use the new features $\tilde{x}(S) = 2x(S)$? Why might this be a problem?

Q2

4. Consider the following general SGD update rule with a general target U_t

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \alpha \left[U_t - \hat{v}(S_t, \mathbf{w}_t) \right] \nabla \hat{v}(S_t, \mathbf{w_t}).$$

Assume we are using linear function approximation, *i.e.* $\hat{v}(S, \mathbf{w}) = x(S)^\top \mathbf{w}$.

(a) What happens to the update if we scale the features by a constant and use the new features $\tilde{x}(S) = 2x(S)$? Why might this be a problem?

- ∇v̂(S,w) will be twice as big, so what does that do to the update…

4. Consider the following general SGD update rule with a general target U_t

Q2

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \alpha \left[U_t - \hat{v}(S_t, \mathbf{w}_t)\right] \nabla \hat{v}(S_t, \mathbf{w_t}).$$

Assume we are using linear function approximation, *i.e.* $\hat{v}(S, \mathbf{w}) = x(S)^\top \mathbf{w}$.

(a) What happens to the update if we scale the features by a constant and use the new features $\tilde{x}(S) = 2x(S)$? Why might this be a problem?

(b) In general we want a stepsize that is invariant to the magnitude of the feature vector $x(S)$, where the magnitude is measured by the inner product $x(S)^\top x(S)$. The book suggests the following stepsize:

$$\alpha = \frac{1}{\tau x(S)^\top x(S)}.$$

What is $x(S)^\top x(S)$ when using tile coding with 10 tilings? Suppose $\tau = 1000$, what is α if we use tile coding with 10 tilings?

Q2

(b) In general we want a stepsize that is invariant to the magnitude of the feature vector $x(S)$, where the magnitude is measured by the inner product $x(S)^\top x(S)$. The book suggests the following stepsize:

$$\alpha = \frac{1}{\tau x(S)^\top x(S)}.$$

What is $x(S)^\top x(S)$ when using tile coding with 10 tilings? Suppose $\tau = 1000$, what is α if we use tile coding with 10 tilings?

- What is x(S)^Tx(S) ??