**Name:** _____

## All Greek to me!

2. (8 points) Let's consider solving a ridge regression problem using stochastic gradient descent. For simplicity, we will ignore the offset. Our hypothesis has the form

$$h(x;\theta) = \theta^T x \ ;$$

our objective function has the form

$$J(\theta) = \left( \frac{1}{n} \sum_{i=1}^{n} \left( h(x^{(i)};\theta) - y^{(i)} \right)^2 \right) + \lambda \|\theta\|^2 \ ;$$

and we will do $T$ steps of gradient descent using a rule of the form

$$\theta = \theta - \eta \nabla_\theta J(\theta) \ ,$$

where $\eta$ has a fixed value throughout the execution.

What is with all these Greek letters!? Each of $\theta$, $\lambda$, and $\eta$ has a role in what happens.

*In the following questions, mark all answers that apply.*

(a) Which parameter(s) would be included when using the hypothesis to make predictions?
◯ $\theta$ ◯ $\lambda$ ◯ $\eta$ ◯ none

(b) Which parameter(s) are primarily intended to improve generalization?
◯ $\theta$ ◯ $\lambda$ ◯ $\eta$ ◯ none

(c) Can $T$ play a similar role to $\lambda$ ?
◯ yes ◯ no
Explain your answer.

(d) Can $\eta$ play a similar role to $\lambda$ ?
◯ yes ◯ no
Explain your answer.

**Name:** _____