1. Give an example of a strongly convex loss function
2. Why do we use gradient descent instead of newtons method even though newtons method converges faster?
3. What is the relationship between the dot product and the Euclidean L2 norm? Is there a practical difference between calculating the norm in these two different ways?

1. The hinge loss function used in SVM
2. Because it is computationally cheaper
3. ||x|| = sqrt{x • x}; Calling np.sqrt(np.dot(x,x)) will be faster than calling np.linalg.norm(x) since np.dot is a highly optimized function that uses BLAS subroutines.