

**Question 1:** What are the differences between Gradient Descent and Newton's method?

**Question 2:** In what cases does gradient descent not work?

**Question 3:** What is the importance of the condition number?

**Answer Question 1:** Both GD and Newton's try to decrease the parameter to a minimum, but GD does it through a learning rate formula and adjusting the weights. Newton's method instead decreases the empirical risk. Newton's also converges faster, as it uses the second derivative, as opposed to GD which uses the first derivative.

**Answer Question 2:** Gradient fails to work in the cases where the learning rate is too large, and when the value of the gradient is zero. When the learning rate is too large, the weights are not guaranteed to decrease and it may fail to converge.

**Answer Question 3:** The condition number indicates how difficult a strongly convex problem is to solve. In gradient descent it refers to how many iterations would need to be done in order for weights to converge.