* Subgradients are used so that we can show that gradient descent can still be used on (some?) nondifferentiable functions