normal equation

In the "Normal Equation" method, we will minimize J by explicitly taking its derivatives with respect to the θj s and setting them to zero. This allows us to find the optimum theta without iteration.

Set the derivative of the cost function to zero as the local optimum’s derivative must be zero.

normal equation vs gradient descent

gradient descent: | normal equation:

need to choose learning rate α | don't need to choose α

need to do many iterations | don't need to iterate - computed in one step

works well with large n | slow if n is large (*n* ⩾ 10000)

| if (*XTX*) is not-invertible - we have problems