Adapted from "Machine Learning: A Probabilistic Perspective" (Murphy), chapter 10