Hyperparameter	Binary Classification	Multiclass Classification
Input Layer Shape	Same as number of features(e.g. 5 for age, sex, height, weight, smoking status, in heart disease prediction)	Same as binary classification
Hidden layer(s)	Problem specific, minimum = 1, maximun = unlimited	Same as binary classification
Neurons per hidden layer	Problem specific, generally 10 to 100	Same as binary classification
Output layer shape	1 (one class or the other)	1 per class (e.g. 3 for food, person or dog photo)
Hidden activation	Usually ReLU (Rectified Linear Unit)	Same as binary classification
Output activation	Sigmoid	Softmax
Loss function	Cross_entropy (tf.keras.losses.BinaryCrossentropy() in TensorFlow	Cross_entropy (tf.keras.losses.CategoricalCrossentropy() in TensorFlow
Optimizer	SGD (Stochsastic Gradient Descent), Adam	Same as binary classification