

## USING SYMCAT SYMPTOM DISTRIBUTION

Assume we have a condition: COVID 19

Assume **ALL** possible symptoms for COVID 19 are **Symptom 1, Symptom 2, ... Symptom k**

Here is the symptom distribution for COVID 19 (this is similar to the dataset we have currently)

COVID 19	Symptom 1	50%
COVID 19	Symptom 2	30%
COVID 19	Symptom 3	60%
COVID 19	Symptom k	20%

Now, lets say we want to create a synthetic training dataset that has **at least 10** training examples **per condition**. How would this training set look like based on the provided symptom distribution?

Patient Name	Symptoms	Symptoms	Symptoms	Symptoms	Diagnosis
Patient1	Symptom 1	Symptom 2	Symptom 3		COVID 19
Patient2				Symptom k	COVID 19
Patient3	Symptom 1		Symptom 3		COVID 19
Patient4				Symptom k	COVID 19
Patient5	Symptom 1		Symptom 3		COVID 19
Patient6		Symptom	Symptom 3		COVID 19
Patient7	Symptom 1				COVID 19
Patient8			Symptom 3		COVID 19
Patient9	Symptom 1				COVID 19
Patient10		Symptom 2	Symptom 3		COVID 19

This covers symptom distribution.

## USING SYMCAT STRONG PREDICTORS

Now assume we want to incorporate details about whether a symptom is a strong predictor or not.

Lets **say symptom 1 and symptom k are strong predictors** for COVID 19 (notice that symptom k is a strong predictor even though it is the least prevalent among patients – 20%)

How would this change the dataset? I'll demonstrate using a before and after snapshot of the above dataset

BEFORE: Using Binary attribute values.

Patient Name	Symptoms	Symptoms	Symptoms	Symptoms	Diagnosis
Patient1	1	1	1	0	COVID 19
Patient2	0	0	0	1	COVID 19
Patient3	1	0	1	0	COVID 19
Patient4	0	0	0	0	COVID 19
Patient5	1	0	1	0	COVID 19
Patient6	0	1	1	0	COVID 19
Patient7	1	0	0	0	COVID 19
Patient8	0	0	1	0	COVID 19
Patient9	1	0	0	0	COVID 19
Patient10	0	1	1	0	COVID 19

AFTER: Using **weighted** vectors to provide higher weights when symptom 1 and symptom k are present.

Patient Name	Symptoms	Symptoms	Symptoms	Symptoms	Diagnosis
Patient1	0.5	0.25	0.25	0	COVID 19
Patient2	0	0	0	1	COVID 19
Patient3	0.75	0	0.25	0	COVID 19
Patient4	0	0	0	1	COVID 19
Patient5	0.75	0	0.25	0	COVID 19
Patient6	0	0.5	0.5	0	COVID 19
Patient7	1	0	0	0	COVID 19
Patient8	0	0	1	0	COVID 19
Patient9	1	0	0	0	COVID 19
Patient10	0	0.5	0.5	0	COVID 19