1) Naïve Bayes for Text Classification

➔ Classes: flower, vegetable, fruit

P(c) = Nc/N

P(w|c) = Count(w,c)/Count(c) + kV

Class probabilities:

P(flower) = 3/8
P(vegetable) = 2/8
P(fruit) = 3/8

For Document 1:

➔ Conditional probabilities: D1: rose, lily, apple, carrot

P(rose|flower) = (6+0.1)/[13+(0.1 * 14)] = 6.1/14.4
P(lily|flower) = (1+0.1)/[13+(0.1 * 14)] = 1.1/14.4
P(apple|flower) = (0+0.1)/[13+(0.1 * 14)] = 0.1/14.4
P(carrot|flower) = (0+0.1)/[13+(0.1 * 14)] = 0.1/14.4

P(rose|vegetable) = (0+0.1)/[8+(0.1 * 14)] = 0.1/9.4
P(lily| vegetable) = (0+0.1)/[8+(0.1 * 14)] = 0.1/9.4
P(apple| vegetable) = (0+0.1)/[8+(0.1 * 14)] = 0.1/9.4
P(carrot| vegetable) = (1+0.1)/[8+(0.1 * 14)] = 1.1/9.4

P(rose|fruit) = (1+0.1)/[14+(0.1 * 14)] = 1.1/15.4
P(lily| fruit) = (1+0.1)/[14+(0.1 * 14)] = 1.1/15.4
P(apple| fruit) = (2+0.1)/[14+(0.1 * 14)] = 2.1/15.4
P(carrot| fruit) = (1+0.1)/[14+(0.1 * 14)] = 1.1/15.4

P(flower|D1) = 3/8 * 6.1/14.4 * 1.1/14.4 * 0.1/14.4 * 0.1/14.4 = 0.00000059
P(vegetable|D1) = 2/8 * 0.1/9.4 * 0.1/9.4 * 0.1/9.4 * 1.1/9.4 = 0.00000004
P(fruit|D1) = 3/8 * 1.1/15.4 * 1.1/15.4 * 1.1/15.4 * 2.1/15.4 = 0.00001864

Therefore, since P(fruit|D1) is the highest for document D1, the most likely class for D1 is **fruit.**

For Document 2:

➔ Conditional probabilities: D2: pea, pea, lotus, grape

P(pea |flower) = (1+0.1)/[13+(0.1 * 14)] = 1.1/14.4
P(lotus |flower) = (0+0.1)/[13+(0.1 * 14)] = 0.1/14.4
P(grape |flower) = (0+0.1)/[13+(0.1 * 14)] = 0.1/14.4

P(pea |vegetable) = (2+0.1)/[8+(0.1 * 14)] = 2.1/9.4
P(lotus | vegetable) = (1+0.1)/[8+(0.1 * 14)] = 1.1/9.4
P(grape | vegetable) = (0+0.1)/[8+(0.1 * 14)] = 0.1/9.4

P(pea |fruit) = (0+0.1)/[14+(0.1 * 14)] = 0.1/15.4
P(lotus | fruit) = (1+0.1)/[14+(0.1 * 14)] = 1.1/15.4
P(grape | fruit) = (2+0.1)/[14+(0.1 * 14)] = 2.1/15.4

P(flower|D1) = 3/8 * 1.1/14.4 * 1.1/14.4 * 0.1/14.4 * 0.1/14.4 = 0.00000011
P(vegetable|D1) = 2/8 * 2.1/9.4 * 2.1/9.4 * 1.1/9.4 * 0.1/9.4 = 0.00001553
P(fruit|D1) = 3/8 * 0.1/15.4 * 0.1/15.4 * 1.1/15.4 * 2.1/15.4 = 0.00000015

Therefore, since P(vegetable|D2) is the highest for document D2, the most likely class for D2 is **vegetable.**