

$$1. \text{ Color} = \{Y, R\}$$

$$\text{Init: } P(B) = \langle \frac{1}{2}, \frac{1}{2} \rangle$$

$$\text{Shape} = \{\text{hole}, \text{solid}\}$$

$$\text{Data: } 2 \text{ YH}$$

$$2 \text{ YS}$$

$$4 \text{ RH}$$

$$2 \text{ RS}$$

B	P(C B)
1	$\langle .3, .7 \rangle$
2	$\langle .7, .3 \rangle$

B	P(S B)
1	$\langle .6, .4 \rangle$
2	$\langle .4, .6 \rangle$

$$\text{E step: } P(B_i | C_i, S_i) = \alpha P(C | B_i) P(S | B_i) P(B_i)$$

$$P(B | C=Y, S=H) = \alpha P(C=Y | B) P(S=H | B) P(B)$$

$$= \alpha \langle .3, .7 \rangle_B \langle .6, .4 \rangle_B \langle .5, .5 \rangle$$

$$= \langle .39, .61 \rangle \text{ (2 examples)}$$

$$P(B | C=Y, S=S) = \alpha \langle .3, .7 \rangle \langle .4, .6 \rangle \langle \frac{1}{2}, \frac{1}{2} \rangle$$

$$= \langle .22, .78 \rangle \text{ (2 examples)}$$

$$P(B | C=R, S=H) = \alpha \langle .7, .3 \rangle \langle .6, .4 \rangle \langle \frac{1}{2}, \frac{1}{2} \rangle$$

$$= \langle .78, .22 \rangle \text{ 4 examples}$$

$$P(B | C=R, S=S) = \alpha \langle .7, .3 \rangle \langle .4, .6 \rangle \langle \frac{1}{2}, \frac{1}{2} \rangle$$

$$= \langle .61, .39 \rangle$$

$$NB_1 = 2 \cdot .39 + 2 \cdot .22 + 4 \cdot .78 + 2 \cdot .61 = 5.56$$

$$NB_2 = 10 - 5.56 = 4.44$$

$$\text{M step:}$$

$$P(B) = \langle .556, .444 \rangle$$

$$P(C=Y | B=1) = \frac{2(.39) + 2(.22)}{5.56} = .22$$

$$P(C=R | B=1) = \frac{4(.78) + 2(.61)}{5.56} = .78$$

$$P(S=H | B=1) = \frac{2(.39) + 4(.78)}{5.56} = .70$$

$$P(S=S | B=1) = \frac{2(.22) + 2(.61)}{5.56} = .3$$

$$P(C=Y | B=2) = \frac{2(.61) + 2(.78)}{4.44} = .63$$

$$P(C=R | B=2) = \frac{4(.22) + 2(.39)}{4.44} = .37$$

$$P(S=H | B=2) = \frac{2(.61) + 4(.22)}{4.44} = .47$$

$$P(S=S | B=2) = \frac{2(.78) + 2(.39)}{4.44} = .53$$

B	P(C B)
1	$\langle .22, .78 \rangle$
2	$\langle .63, .37 \rangle$

B	P(S B)
1	$\langle .7, .3 \rangle$
2	$\langle .47, .53 \rangle$

E step:

$$P(BIC=Y, S=H) = \alpha \langle .22, .63 \rangle \langle .7, .47 \rangle \langle .556, .444 \rangle$$

$$= \alpha \langle .15, .3 \rangle \langle .556, .444 \rangle$$

$$= \alpha \langle .08, .13 \rangle$$

$$= \langle .38, .62 \rangle \quad 2 \text{ examples}$$

$$P(BIC=Y, S=S) = \alpha \langle .22, .63 \rangle \langle .3, .53 \rangle \langle .556, .444 \rangle$$

$$= \alpha \langle .04, .15 \rangle$$

$$= \langle .21, .79 \rangle \quad 2 \text{ examples}$$

$$P(BIC=R, S=H) = \alpha \langle .78, .37 \rangle \langle .7, .47 \rangle \langle .556, .444 \rangle$$

$$= \alpha \langle .3, .08 \rangle$$

$$= \langle .78, .22 \rangle \quad 4 \text{ examples}$$

$$P(BIC=R, S=S) = \alpha \langle .78, .37 \rangle \langle .3, .53 \rangle \langle .556, .444 \rangle$$

$$= \alpha \langle .13, .09 \rangle$$

$$= \langle .59, .41 \rangle \quad 2 \text{ examples}$$

$$N_{B_1} = 2(.38) + 2(.21) + 4(.78) + 2(.59) = 5.48$$

$$N_{B_2} = 2(.62) + 2(.79) + 4(.22) + 2(.41) = 4.52$$

E step:

$$P(BIC=Y, S=H) = \alpha \langle .15, .3 \rangle \langle .548, .452 \rangle$$

$$= \alpha \langle .08, .13 \rangle = \langle .38, .62 \rangle$$

$$P(BIC=Y, S=S) = \alpha \langle .04, .15 \rangle$$

$$= \langle .21, .79 \rangle$$

$$P(BIC=R, S=H) = \alpha \langle .3, .08 \rangle = \langle .78, .22 \rangle$$

$$P(BIC=R, S=S) = \alpha \langle .13, .09 \rangle = \langle .59, .41 \rangle$$

$$N_{B_1} = 5.48$$

$$N_{B_2} = 4.52$$

No change!

M step:

$$P(B) = \langle .548, .452 \rangle$$

$$P(C|B=1) = \frac{2(.38) + 2(.21)}{5.48}$$

$$\frac{4(.78) + 2(.59)}{5.48}$$

$$= \langle .22, .78 \rangle$$

$$P(S|B=1) = \frac{2(.78) + 4(.78)}{5.48}$$

$$\frac{2(.21) + 2(.59)}{5.48}$$

$$= \langle .7, .3 \rangle$$

$$P(C|B=2) = \frac{2(.62) + 2(.79)}{4.52}$$

$$\frac{4(.22) + 2(.41)}{4.52}$$

$$= \langle .63, .37 \rangle$$

$$P(S|B=2) = \frac{2(.62) + 4(.22)}{4.52}$$

$$\frac{2(.79) + 2(.41)}{4.52}$$

$$= \langle .47, .53 \rangle$$

M step:

$$P(C|B=1) = \langle .22, .78 \rangle$$

$$P(S|B=1) = \langle .7, .3 \rangle$$

$$P(C|B=2) = \langle .63, .37 \rangle$$

$$P(S|B=2) = \langle .47, .53 \rangle \quad \text{No change!}$$

2b. My Classifier seems to Perform very Poorly on the 'iy' and 'ey' vowels. Pretty much all the data centered about the (2500, 400) point. I believe there is just too much data stacked in this area of the Plot.