Homework 1

Jia Jiun Ku
jk2mf

September 22, 2019

Homework 1

1. Using Bayes' theorem, we have

$$P(y \mid x) = \frac{P(x,y)}{p(x)} = \frac{P(x \mid y)P(y)}{P(x)}$$

Consider the cases where $X = 0$:

$$P(Y = 0 \mid X = 0) = \frac{\alpha \cdot 0.5}{0.5\alpha + 0.5(1-\beta)} = \frac{\alpha}{1+\alpha-\beta}$$

$$P(Y = 1 \mid X = 0) = \frac{(1-\beta) \cdot 0.5}{0.5\alpha + 0.5(1-\beta)} = \frac{1-\beta}{1+\alpha-\beta}$$

Since $\alpha > 1 - \beta$, our Naive Bayes classifier will predict $Y = 0$ when the input is $X = 1$.

Now consider when $X = 1$:

$$P(Y = 0 \mid X = 1) = \frac{(1-\alpha) \cdot 0.5}{0.5(1-\alpha) + 0.5\beta} = \frac{1-\alpha}{1-\alpha+\beta}$$

$$P(Y = 1 \mid X = 1) = \frac{\beta \cdot 0.5}{0.5(1-\alpha) + 0.5\beta} = \frac{\beta}{1-\alpha+\beta}$$

$\beta > 1 - \alpha$, so the prediction will be $Y = 1$ when the input is $X = 0$

From the above equations, we can infer the probability of the classifier making an error is

$$\begin{aligned}(X = 0, Y = 1) + P(X = 1, Y = 0) &= P(X = 0 \mid Y = 1)P(Y = 1) \\ &\quad + P(X = 1 \mid Y = 0)P(Y = 0) \\ &= 0.5(1-\beta) + 0.5(1-\alpha)\end{aligned} \tag{1}$$

2. WLOG, assume $\theta_j^{(1)} > \theta_j^{(2)}$. This means training on D_1 gives us a $\theta_j^{(1)} > \theta_j^{(2)}$, meaning D_2 has data that will make the value of θ_j smaller, that data being $D_2 - D_1$. Training on $D_1 \cup D_2$ is essentially training on D_1 and then training on $D_2 - D_1$, which will give us a $\theta_j^* \leq \theta_j^{(1)}$. $\theta_j^* \geq \theta_j^{(2)}$ can be proved in the same way.

3. Suppose towards a contradiction that $\|\theta^*\|_2^2 > \left\|\hat{\theta}\right\|$. By definition, we have that

$$\hat{\theta} = \arg\min_{\theta} J(\theta)$$

and

$$\theta^* = \arg\min_{\theta} J(\theta) + \|\theta\|_2^2$$

and that

$$J(\hat{\theta}) \leq J(\theta^*)$$

Based on our assumption, we have

$$J(\hat{\theta}) + \left\|\hat{\theta}\right\|_2^2 < J(\theta^*) + \|\theta^*\|_2^2$$

. This contradicts the definition of θ^*. Therefore our assumption must be false.

4. Since $(p - r)^2 \geq 0$ we have

$$p^2 + r^2 - 2pr \geq 0$$
$$p^2 + r^2 + 2pr \geq 4pr$$
$$(p + r)^2 \geq 4pr$$
$$\frac{1}{2}(p + r) \geq \frac{2pr}{p + r}$$

So the arithmetic mean of precision and recall is always greater than or equal to F-measure.

When $p = r$, F-measure=

$$\frac{2pr}{p + r} = \frac{2p^2}{2p} = \frac{2p}{2} = \frac{p + r}{2}$$

When $F = \frac{p+r}{2}$,

$$\frac{2pr}{p + r} = \frac{p + r}{2}$$
$$\frac{4pr}{p + r} = p + r$$
$$4pr = (p + r)^2$$
$$0 = p^2 - 2pr + r^2 = (p - r)^2$$
$$=> p = r$$