4. Теоретические задачи.

1.

$$a(x) = \arg\max_y P(y) \prod_{k=1}^{n} P(x^{(k)} | y)$$

$$P(x^{(k)} | y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x^{(k)} - \mu_{yk})^2}{2\sigma^2}}$$

$$a(x) = \arg\max_y P(y) \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^{(k)} - \mu_{yk})^2}{2\sigma^2}} = \left| \frac{1}{\sqrt{2\pi\sigma^2}} - \text{const} \right|$$

$$= \arg\max_y P(y) \prod_{k=1}^{n} e^{\frac{-(x^{(k)} - \mu_{yk})^2}{2\sigma^2}} = \left| P(y) - \text{априорная} \right| =$$

$$= \arg\max_y \sum_{k=1}^{n} -(x^{(k)} - \mu_{yk})^2 = \arg\min_y \sum (x^{(k)} - \mu_{yk})^2.$$

Таким образом, классификация сводится к отнесению объекта x к классу y, центр которого μ_y ближе всего к x.

2.

$a(x) = 1 \; : \; p$

$a(x) = 0 \; : \; 1-p$

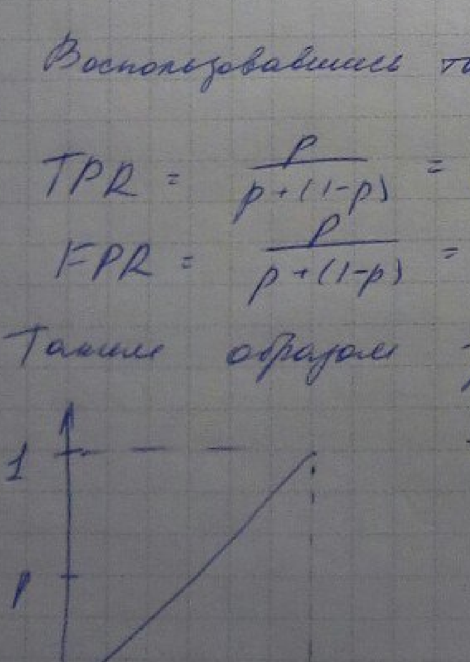
		Actual Class	
		Yes	No
Predicted Class	Yes	p	p
	No	$1-p$	$1-p$

Воспользовавшись таблицей, подсчитаем TPR и FPR:

$$TPR = \frac{p}{p + (1-p)} = p$$

$$FPR = \frac{p}{p + (1-p)} = p$$

Таким образом график будет представлять из себя прямую.



Площадь под графиком не будет зависеть от p и будет равна $\frac{1^2}{2} = \frac{1}{2}$

$$\Rightarrow E(AUC) = \frac{1}{2},$$

независимо от p и доли класса в обучающей выборке.