

1. **Page 15: testing set**

All those data are in my computer at the campus. They will be uploaded when I am back to school. Right now, only the codes and the results are available in my github repository.

2. **Page 25: Should we compare this with the non-core region**

Done

3. **Page 32: I do not think this sentence is correct. AUC has different meaning.**

The interpretation of the AUC values were deleted. But, I think, to some extent, we can interpret the AUC as the probability to distinguish a randomly chosen positive sample from a randomly chosen negative sample. The arguments are as follows.

Suppose M is the model, x is the input, T is the threshold. If $M(x) > T$, x is classified as positive. Otherwise, x is classified as negative. FPR and TPR can be viewed as two functions of T , given by

$$\begin{aligned} TPR(T) &= \int_T^\infty f_0(y)dy \\ FPR(T) &= \int_T^\infty f_1(y)dy \end{aligned}$$

Here, f_0 is the distribution density of the set $\{M(x) : x \text{ is a positive sample}\}$, and f_1 is the distribution density of the set $\{M(x) : x \text{ is a negative sample}\}$. The AUC is given by the following equations

$$\begin{aligned} AUC &= - \int_{-\infty}^\infty TPR(T) d(FPR(T)) \\ &= \int_{-\infty}^{+\infty} TPR(T) f_1(T) dT \\ &= \int_{-\infty}^{+\infty} \int_T^\infty f_0(y) f_1(T) dy dT \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^\infty I(y > T) f_0(y) f_1(T) dy dT \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^\infty I(y_0 > y_1) f_0(y_0) f_1(y_1) dy_0 dy_1 \\ &= P(y_0 > y_1) \\ &= P(M(x_0) > M(x_1)) \end{aligned}$$

Here, $y_0 = M(x_0)$, $y_1 = M(x_1)$, where x_0 is a randomly chosen positive sample and x_1 is a randomly chosen negative sample.

4. **Page 33: Are there methods from other investigators to achieve this task? If yes, you should compare with them. I do not know the field well enough to give references and which methods to compare with. But my gut feeling is that there should be some you can compare with.**

As far as I know, there is no research specifically focused on this task. There are researches slightly related to this task, as introduced in the introduction chapter, but they are aiming to predict epitopes. To some extent, this task is new. The method we used is new as well.

5. **Page 49: I cannot see data for this? Did you provide data to support this claim?**

Table 4.3 on page 45 can give some support for this claim.