Lecture 15 - 7: Prediction

Tabaré Pérez

May 6, 2020

So now, we are ready to start looking at the question of prediction. So again, as the same way as in the case of our discriminative supervised model, we will have our points, let's say, just two classes, pluses and minuses.

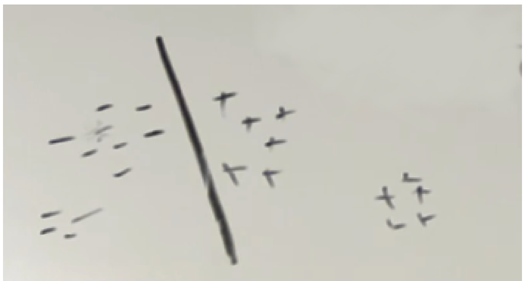

Figure 1: Discriminative supervised model

And using the estimation techniques, as I just described to you earlier, we can actually induce the probability distribution: find θs.

Find θs for plus points, θ^+, and also, find θs for minus points, θ^-.

Find the parameters for thes classes in such a way that they will give the highest likelihood to the points, for instance, on the plus side.

So now, the question is, if I give you a new document, how do you know to which class it belongs?

So one way to think about it is to say, I have my document and I could look at the likelihood that this document was generated by plus side:

$$\mathbb{P}(D|\theta^+) \tag{1}$$

Also I could look at the likelihood that this document is generated by the minus class:

$$\mathbb{P}(D|\theta^-) \tag{2}$$

And look its log:

$$\log\left(\frac{\mathbb{P}(D|\theta^{+})}{\mathbb{P}(D|\theta^{-})}\right) \tag{3}$$

You can ask me why I am looking at the log. You will see later.

But the point is, intuitively speaking, we would want to see that the document will be assigned to the class which gives to it the higher likelihood. And for now, I will make one assumption, which I am going to break later. Let's make an assumption that the likelihood of a document in the minus class (the prior likelihood of the minus class) and in the plus class are exactly the same.

So in this particular case, we are going to be looking at this log:

$$\log\left(\frac{\mathbb{P}(D|\theta^{+})}{\mathbb{P}(D|\theta^{-})}\right)\begin{cases} +, & if \geq 0 \\ -, & otherwise \end{cases} \tag{4}$$

And we would say that if it is bigger or equal to zero, then we are going to return plus. and if it is smaller than zero, we are going to return the minus.

So these type of distributions are called **class conditional distributions** so that you just know.

So let's just look more closely. I am going to just open up this expression and massage it in a different way. And eventually, I will bring it to you to the form that you've already seen in the past. But let's just start. So again, we have here log of this ratio and we can just write it as log of one expression minus the log of another expression:

$$\log\mathbb{P}(D|\theta^{+}) - \log\mathbb{P}(D|\theta^{-}) = \log\prod_{w\in\mathcal{W}}\theta^{+^{\text{count}(w)}} - \log\prod_{w\in\mathcal{W}}\theta^{-^{\text{count}(w)}} = \tag{5}$$

So now, we can continue to do some manipulation. and again, we'll remember that log of the product is the sum of logs:

$$= \sum_{w\in\mathcal{W}}\text{count}(w)\cdot\log(\theta_w^{+}) - \sum_{w\in\mathcal{W}}\text{count}(w)\cdot\log(\theta_w^{-}) = \tag{6}$$

So I will do my almost last rearrangement here:

$$= \sum_{w\in\mathcal{W}}\text{count}(w)\cdot\overbrace{\log\left(\frac{\theta_w^{+}}{\theta_w^{-}}\right)}^{\hat{\theta}_w} = \tag{7}$$

And the last thing that I'm going to do, this is truly the last thing, I am going to just use to introduce for you and a new notation.

So instead of writing this big expression, , I'm going to just call it like $\hat{\theta}_w$. So I'm just writing it here:

$$= \sum_{w \in \mathcal{W}} \text{count}(w) \cdot \hat{\theta}_w \tag{8}$$

And you would immediately see why it is a good idea.

So if you are now looking at this, so remember we started here:

$$\log \left(\frac{\mathbb{P}(D|\theta^+)}{\mathbb{P}(D|\theta^-)} \right) = \sum_{w \in \mathcal{W}} \text{count}(w) \cdot \hat{\theta}_w \tag{9}$$

So at this point, looking at this expression, which we derived looking through generative view on classification, actually what we got here should remind you a linear classifier that goes through origin with respect to this parameter $\hat{\theta}_w$. So despite the fact that we went kind of in a very different way, what we got with our generative model, is a linear classifier, just get there in a different way.