CIS399: The Science of Data Ethics	Lecture 23: April 11, 2019
Instructor: Michael Kearns & Ani Nenkova	Scribes: Ajani Motta, Kenan Saleh

1 Data Utilization Tradeoffs

Last class we talked about the tradeoff between the privacy concerns of using personal data with the beneficial effects of using it to make decision-making more accurate.

One such example is in medicine, where we can use individuals' medical data to have strong predictive power that can be potentially life-saving, yet we may need to access sensitive behavioral information of individuals (such as their Google search history and medical records) in order to do so. Another application in medicine is to determine which medication to administer to patients with more accuracy by looking at their genome and other medical indicators (particularly for drugs that are very patient-specific such as cancer treatment). Theres potential for this process to transform medicine to be more personalized and effective. Though it may be scary to make this kind of data available to many parties (which is a valid concern), it clearly could have and is having very beneficial results. Finding the right balance of this tradeoff is tricky. Another interesting aspect of data in medicine is that it is impossible to be race-blind or gender-blind since there are many diseases that have very different probabilities of occurring and treatments based on these factors.

In previous classes, we presented a technique of embedding our desired fairness constraints into our classifier models. So have we solved the accuracy vs fairness tradeoff? Not quite. Weve shown how to make tradeoffs with our models but its still not clear which tradeoffs should be made and which are actually useful. We will present such a framework in todays class.

2 Net Utility

Lets start by looking at Net Utility. Net Utility is the net benefit achieved by taking a certain action or tradeoff. Lets consider a diagnostic test. We will use the True Positive Rate (TPR) and False Positive Rate (FPR) metrics we've discussed in prior classes to evaluate Net Utility:

$$TPR = \frac{TP}{TP + FP} \tag{1}$$

$$FPR = \frac{FP}{TP + FP} \tag{2}$$

If we plot False Positive Rate vs True Positive Rate and vary the confidence threshold (the probability for which the model will predict x to be in the positive class), we get a curve with each point representing a different possible classifier within this family of classifiers. The question is: what classifier should I choose? Which provides the most optimal tradeoff in terms of Net Utility?

Lets define Net Utility to be: $A * (\frac{TP}{n}) + B * (\frac{FP}{n})$ where n represents the number of items in the test set, A represents the benefit of a true positive, and B represents the penalty of a false positive. How do we determine A and B? We will set $A = 1$ and $B = \frac{-pt}{1-pt}$ where pt is the threshold probability weve used for a classifier. Now we can try to optimize for a single metric: Net Utility. Now we can choose the threshold probability and corresponding classifier that leads to the highest Net Utility.

3 Decision Curve

Decision curves measure net utility as a function of threshold probability (p_t) of a particular model. It operates under the central assumption that this threshold probability (at which a patient would opt for treatment) is important in how a patient determines the value of a false positive or false negative prediction. These curves work to facilitate the process of finding an "optimal" balance between the accuracy and fairness of a family of models.

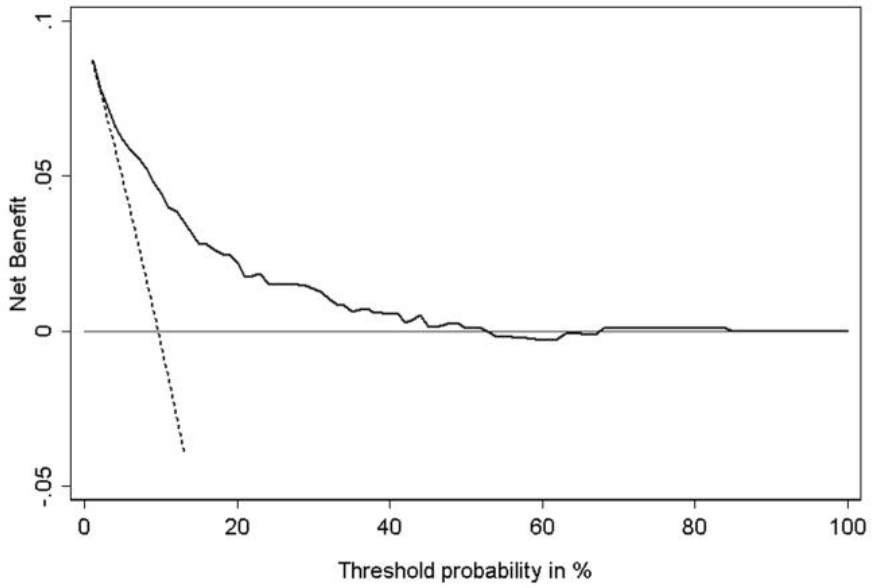

Figure 1: Decision curve for a model to predict seminal vesicle invasion (SVI) in patients with prostate cancer. The dark curve represents the net benefit achieved by the prediction model. The dotted line represents the net benefit achieved by treating all patients. The gray line represents the net benefit achieved by treating no patients.

These curves are often used to analyze the performance of a model running according to different threshold probabilities. Using the net benefit function along with the decision curve allows decision makers to quantify changes in bias across different thresholds. Figure 2 provides an example of how this can be done.

4 Drawbacks

Due to the fluid nature of the fairness problem, there is no "universal" solution; no one way in which a model can be treated to achieve an optimal balance between accuracy and fairness. Although the net benefit paradigm provides a framework for assessing this balance, it still provides metrics that aim to quantify unquantifiable notions (i.e. harm of false positive and false negative predictions). There is almost always an argument that can be made for applying different weights on such harms and how fairness is defined.

p_t (%)	Net Benefit		Advantage of model	
	Treat All	Prediction Model	Net benefit	Reduction in avoidable tip surgeries per 100 patients
1	0.087	0.087	0	0
2	0.078	0.078	0	0
3	0.069	0.072	0.004	13
4	0.059	0.066	0.007	17
5	0.049	0.062	0.013	25
6	0.039	0.059	0.020	31
7	0.028	0.056	0.027	36
8	0.018	0.053	0.035	40
9	0.007	0.048	0.041	41
10	−0.004	0.044	0.048	43

Figure 2: Decision table. The last column is the result of:
$(\frac{1-p_t}{p_t}) * (\text{net benefit of model - net benefit of treat all})$.
$\Longrightarrow$ the net value of false negatives

4.1 Reasoning behind structure of net utility function

Consider the case where a patient has to decide whether to undergo treatment for a particular disease, where the patient is unsure as to whether she/he has the disease. Aligning with the decision tree below, p = probability of the disease, $\{a, b, c, d\}$ represent the utility corresponding to each outcome. A prediction model provides the probability p, where the patient is likely to undergo treatment if p is close to 1 and contrarily if p is close to 0. This implies that there is some point between 0 and 1, where the patient will show indifference between the two choices. This probability is considered the threshold probability (p_t), where the utility gained is equal (treatment vs. no treatment). The conditions at the threshold probability imply:

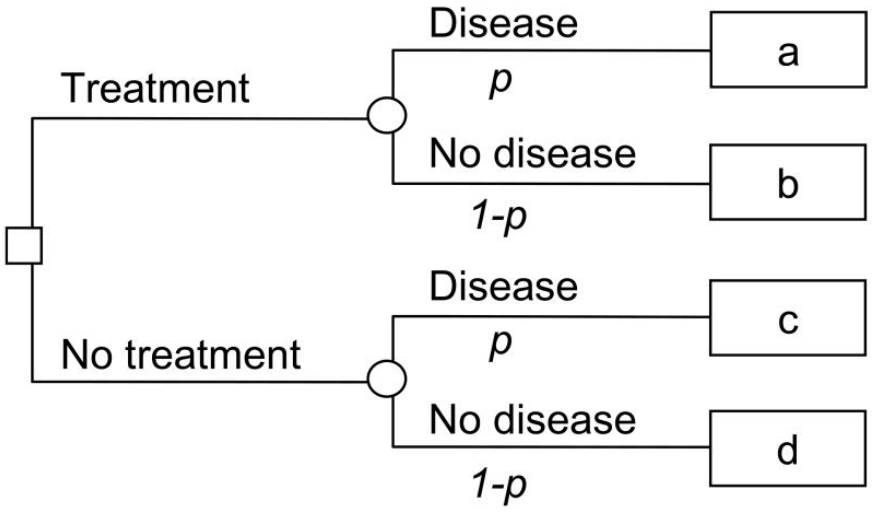

Figure 3: Decision tree

$$Utility(treatment) = Utility(!treatment) \tag{3}$$

$$(p_t)a + (1 - p_t)b = (p_t)c + (1 - p_t)d \tag{4}$$

$$(a - c)p_t = (d - b)(1 - p_t) \tag{5}$$

$$\frac{b - d}{a - c} = \frac{-p_t}{1 - p_t} \tag{6}$$

This shows us that the threshold probability at which a patient opts for treatment is important in how a patient weighs the harm of a false positive relative to a false negative.