Data Mining	Fall 2021
Tu, Th 2:00-3:15	HW 3 UG

Problem 1 [4pts]-1pt per feature . Given the data set fig. 1 that is associated to a node in a tree. Should this node be split if we are using χ^2 pre-pruning aproach at significance lebel 0.05? on which feature? Consider only multi-outcome splits. (grading: 1pt per test)

Age	Spectacle Prescription	Astigmatism	Tear Production Rate	Recommended Contact Lens
Young	Myope	No	Reduced	None
Young	Myope	No	Normal	Soft
Young	Myope	Yes	Reduced	None
Young	Myope	Yes	Normal	Hard
Young	Hypermetrope	No	Reduced	None
Young	Hypermetrope	No	Normal	Soft
Young	Hypermetrope	Yes	Reduced	None
Young	Hypermetrope	Yes	Normal	Hard
Pre-presbyopic	Myope	No	Reduced	None
Pre-presbyopic	Myope	No	Normal	Soft
Pre-presbyopic	Myope	Yes	Reduced	None
Pre-presbyopic	Myope	Yes	Normal	Hard
Pre-presbyopic	Hypermetrope	No	Reduced	None
Pre-presbyopic	Hypermetrope	No	Normal	Soft
Pre-presbyopic	Hypermetrope	Yes	Reduced	None
Pre-presbyopic	Hypermetrope	Yes	Normal	None
Presbyopic	Myope	No	Reduced	None
Presbyopic	Myope	No	Normal	None
Presbyopic	Myope	Yes	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	No	Reduced	None
Presbyopic	Hypermetrope	No	Normal	Soft
Presbyopic	Hypermetrope	Yes	Reduced	None
Presbyopic	Hypermetrope	Yes	Normal	None

Figure 1: Lenses data set

Problem 2 [5pts] . Given the data set fig. 1, the following tree (fig. 2was constructed.

[1pt] Is this tree overfitted? if so explain why, if not give your argument why not.

[4pts] Can you tell if at 0.05 confidence level the node 'age [N=1,S=5,H=0]' is useful or is this split unnecessary?

- Use $e \leq \max\{e_i\}_{i \in \{\texttt{Pr},\texttt{Y},\texttt{Pre-Pr}\}}$
- Treat non-leaf node as if it was a leaf (i.e. classify its data by majorityand use respective empiric error)

  Hint: recall that confidence level has to be recomputed for the tree even when we use crude estimate f routing probability

FYI credit structure: 1pt per node + 1pt per tree)

Bonus

Let X be a discrete domain, and let $\mathcal{H}_{\text{Singleton}} = \{h_z : z \in X\} \cup \{h^-\}$, where for each $z \in X$, h_z is the function defined by $h_z(x) = \begin{cases} 1 \text{ if } x = z \\ 0 \text{ otherwise} \end{cases}$ and h^- is simply the all-negative hypothesis, namely, $\forall x \in X, h^-(x) = 0$. The realizability assumption here implies that the true hypothesis f labels negatively all examples in the domain, perhaps except one.

[5pts] Describe an algorithm that implements the ERM rule for learning $\mathcal{H}_{\text{Singleton}}$ in the realizable setup.

[6pts] Show that $\mathcal{H}_{\text{Singleton}}$ is PAC learnable. Provide an upper bound on the sample complexity

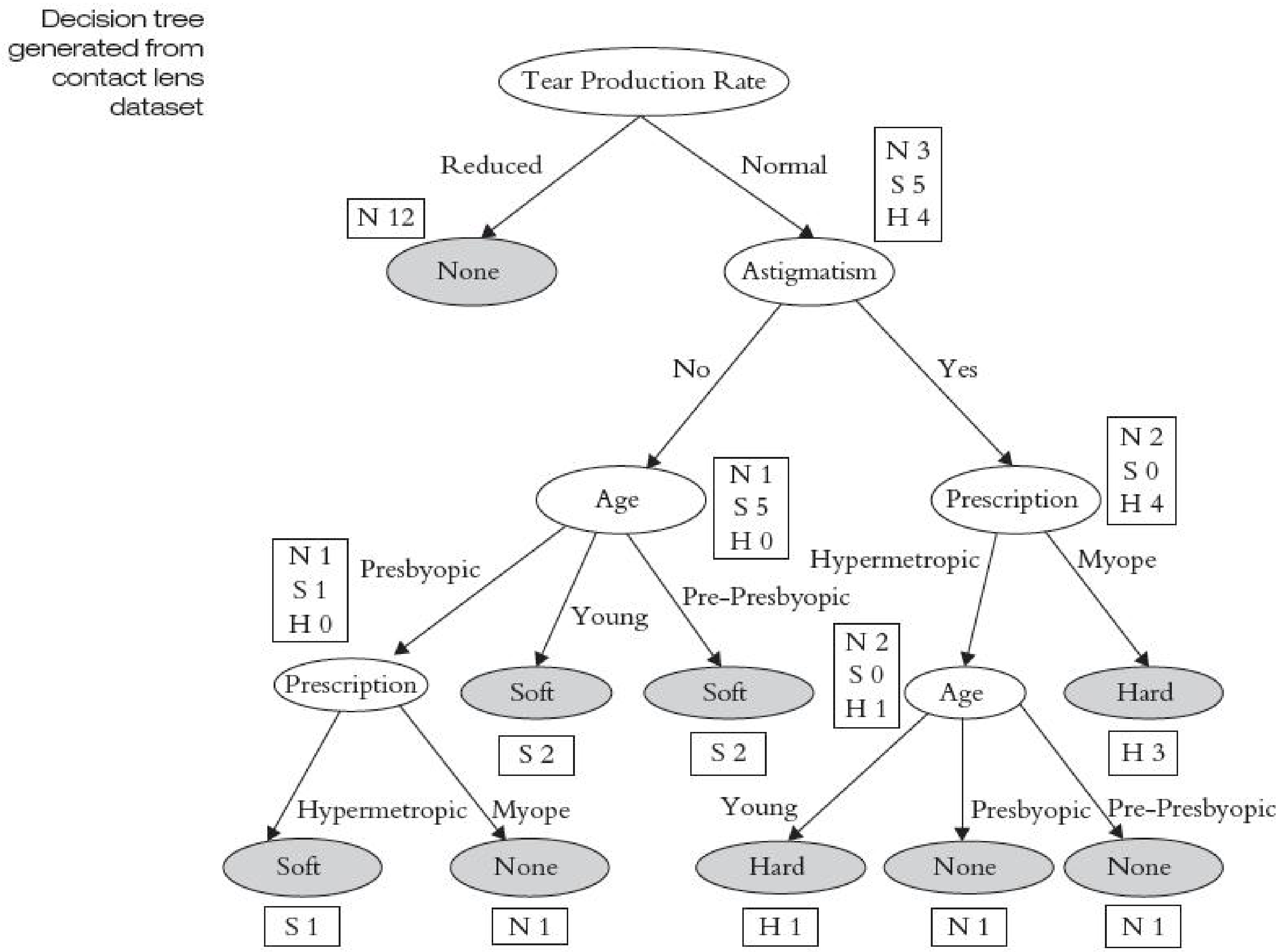

Figure 2: Decision Tree for the Lenses data set