

## 人工知能統論課題 9

### [課題 9]

以下のコンピュータに関するトレーニングデータの集合から、各コンピュータ ( $x_1$  から  $x_8$  まで) が「上位モデル」に属するか否かを判別する決定木を生成する問題を考える。以下において、対数の底は 2 で、計算を簡単にするために  $\log_2 3 = 1.5$  として計算しなさい。答えは規約分数として求めよ。なお、 $\log \frac{X}{Y} = \log X - \log Y$  であることに注意。

トレーニング データ	属 性			上 位 モデル
	Price	Clock	Design	
$x_1$	high	slow	good	No
$x_2$	high	fast	moderate	Yes
$x_3$	high	middle	good	Yes
$x_4$	high	middle	moderate	Yes
$x_5$	low	middle	good	No
$x_6$	low	fast	bad	Yes
$x_7$	low	slow	moderate	No
$x_8$	low	slow	bad	No

#### [課題 9-1] (10 点)

このトレーニングデータの集合のエントロピ (平均情報量) を求めなさい。

#### [課題 9-2](30 点)

このトレーニングデータの集合に対して”Price”, ”Clock”, ”Design” の各属性テストを適用した場合、それぞれのテストに対するエントロピのゲイン (利得) を求めよ。この場合のゲインとは、テスト前の集合のエントロピ (課題 9-1 で求めた値) とテスト後の全データ集合の平均エントロピの差である。

#### [課題 9-3](10 点)

このトレーニングデータに ID3 を適用した場合の決定木を求めよ。