ITM885 Homework 1
Decision making, optimization and penalized regression

Due on Thursday 09/23 before class (100 points)

1 Freemark Abbey Winery (25 points)

Read *FreemarkAbbeyWinery.pdf* carefully before answering the following questions. You should assume that Freemark Abbey Winery sells the wine in bulk ($1 per bottle) if the storm hits and there is no mold.

1. Fill the payoff table below using the information given in the case. Find the optimal action with the *maximin* rule, the optimal action with the *maximax* rule.

Table 1: Payoff Table

	0.50*0.40 =0.20 Storm Botrytis	0.50*0.60 =0.30 Storm No Botrytis	0.50*0.40 =0.20 No Storm Sugar 25%	0.50*0.40 =0.20 No Storm Sugre 20%	0.50*0.20 =0.10 No Storm Acidity <0.7%
Harvest Now	2.85*12 =34.2	2.85*12 =34.2	2.85*12 =34.2	2.85*12 =34.2	2.85*12 =34.2
Harvest Later	8*12*0.7 =67.2	2*12/2 =12	3.5*12 =42	3.0*12 =36	2.5*12 =30

maximin: Harvest Now

maximax: Harvest Later

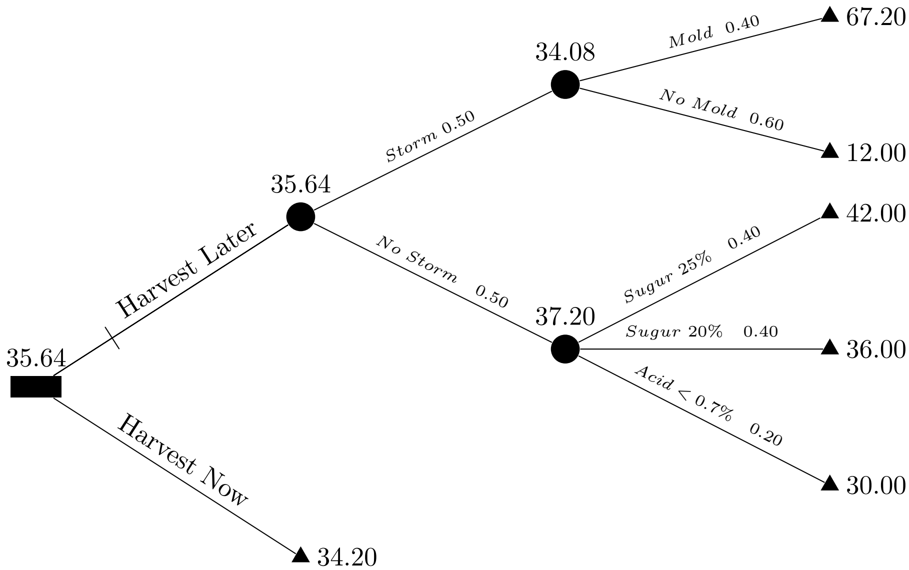
2. Create a loss table and find the optimal action with the *minimax loss* criterion.

   minimax loss: Harvest Later

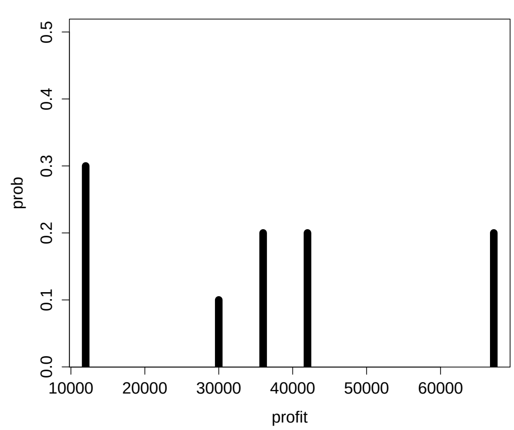
3. Construct a decision tree using the information given in the case.

Table 2: Loss Table

	0.50*0.40 =0.20 Storm Botrytis	0.50*0.60 =0.30 Storm No Botrytis	0.50*0.40 =0.20 No Storm Sugar 25%	0.50*0.40 =0.20 No Storm Sugre 20%	0.50*0.20 =0.10 No Storm Acidity <0.7%
Harvest Now	33	0	7.8	1.8	0
Harvest Later	0	22.2	0	0	4.2



4. What is the probability distribution that represents the uncertainty regarding the possible outcomes if Jaeger decides to wait to see if the storm hits (rather than harvest immediately)? What is the mean of this distribution?



The mean is \$35.6 K.

5. What decision would you recommend to Jaeger given the information you have?

   Harvest Later to maximize the expected payoff.

6. Would your decision change if the probability changes from 0.4 to 0.2 that the botrytis mold forms given that the storm hits? Why or why not?

   Yes, because the expected payoff of Harvest Later changes to \$30.1 K, smaller than \$34.2 K. So the new recommendation that maximizes the expected payoff would be Harvest Now.

7. Suppose Jaeger's utility function for x thousand dollars is

$$U(x) = 1 - e^{-\frac{x}{100}}.$$

   Find the optimal action in terms of expected utility.

   The expected utility for Harvest Now is $1 - e^{-34.2/100} = 0.290$.

   The expected utility for Harvest Later is $(1 - e^{-67.2/100}) * 0.2 + (1 - e^{-12/100}) * 0.3 + (1 - e^{-42/100}) * 0.2 + (1 - e^{-36/100}) * 0.2 + (1 - e^{-30/100}) * 0.1 = 0.287$.

   Thus the optimal action in terms of expected utility is Harvest Now.

2 Ridge regression (10 points)

For a linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, ridge regression has the loss function as MSE plus an L_2 penalty. Specifically, $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda ||\boldsymbol{\beta}||_2^2$, where $\lambda > 0$ is a hyper-parameter to be tuned. First prove the optimization problem is convex. Then given λ, derive $\hat{\boldsymbol{\beta}}$ by the matrix differentiation that is analogous to that of OLS. Does $\hat{\boldsymbol{\beta}}$ always exist?

Solution:

The optimization problem is convex because both terms in the objective function are quadratic forms. Since $||\boldsymbol{\beta}||_2^2 = \boldsymbol{\beta}'\boldsymbol{\beta}$, we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \arg\min_{\boldsymbol{\beta}} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda |\boldsymbol{\beta}|^2 \\
&= \arg\min_{\boldsymbol{\beta}} -2\boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{I})\boldsymbol{\beta}.
\end{aligned}$$

Setting the gradient of the objective function to $\mathbf{0}$, we have $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}$ which always exists as $(\boldsymbol{X}'\boldsymbol{X} + \lambda \boldsymbol{I})$ is positive definite and invertible.

3 Penalized regression (30 points)

Use `PyTorch` or the `optim` function in `R` to fit a ridge regression by gradient descent for the prostate cancer data. You can choose a set of grid values for λ and select the best λ by cross-validation and the minimum CV rule. Report the best λ, the corresponding estimated regression coefficients, and RSS_{tst}.

Solution:

The result depends on the condidate values of λ, how the data is preprocessed (whether centralized or normalized or not) and shuffled for cross-validation.