Lectures Notes #2 Jan 23

A Game against Nature

Let Θ = a set of choices for nature. $\theta \in \Theta$ is a particular choice for nature, and $u \in U$ is our choice. The Loss function again is $L : U \times \theta \to \Re$, $L(u, \theta)$
For example, consider the following table

$$\begin{bmatrix} & u_1 & u_2 & u_3 \\ \theta_1 & 1 & -1 & 2 \\ \theta_2 & -1 & 2 & -1 \\ \theta_3 & 0 & -2 & 1 \end{bmatrix}$$

Two ways to make a decision
Nondeterministic: I have no idea - use Murphy's Law to make a decision (assume the worst case scenario). In our example we would choose u_1.
Probabilistic: Observe nature over many trials and use Baysian analysis to make decisions.

Suppose that $P(\theta)$ is known (informed decision making).
In the **Nondeterministic** version then we would choose u to minimize

$$max_{\theta \in \Theta} L(u, \theta)$$

This is a min-max solution.
In the **Probabilistic** version we would choose u to minimize

$$E^{\theta}(L) = \sum_{\theta \in \Theta} L(u, \theta) \cdot P(\theta)$$

Laplace in (1812) said to simply choose the column with the lowest average loss. He assumed that each $\theta \in \Theta$ is equally likelly.

Minumum Doh! (regret)
Choose u_i such that it will bring you the least regret. The following table is based on the previous one that we had.

$$\begin{bmatrix} & u_1 & u_2 & u_3 \\ \theta_1 & 2 & 0 & 3 \\ \theta_2 & 0 & 3 & 0 \\ \theta_3 & 2 & 0 & 3 \end{bmatrix}$$

The best choice for this case is u_1. This is completely coinsidental that u_1 is the best solution for both matrices. If we would have had originally the following matrix:

$$\begin{bmatrix} & u_1 & u_2 & u_3 \\ \theta_1 & 1 & -1000 & 2 \\ \theta_2 & -1 & 2 & -1 \\ \theta_3 & 0 & -2 & 1 \end{bmatrix}$$

Note that u_1 is still the best choice here. Then the regret matrix would have been

$$\begin{bmatrix} & u_1 & u_2 & u_3 \\ \theta_1 & 1001 & 0 & 1002 \\ \theta_2 & 0 & 3 & 0 \\ \theta_3 & 2 & 0 & 3 \end{bmatrix}$$

However in order to minimize the regret the best choice would be $u_2 \neq u_1$. The probabilistic version here would be to choose u_i such that the expected regret is minimized.

Now lets look at a couple scenarios
Scenario 2 Min-max solution, nondeterministic
1. $U = \{u_1, \ldots, u_n\}$
2. $\Theta = \{\theta_1, \ldots, \theta_m\}$
3. $L : U \times \Theta \rightarrow \Re$
4. Choose u to minimize $max_{\theta \in Theta} L(u, \theta)$
Note that this scenario is very pessimistic!! The complete opposite would be to hava a very optimistic policy.

Scenario 3 Expected Optimal Solution
1. $U = \{u_1, \ldots, u_n\}$
2. $\Theta = \{\theta_1, \ldots, \theta_m\}$
3. $P(\theta)$ given $\forall\, \theta \in \Theta$
4. $L : U \times \Theta \rightarrow \Re$
5. Choose u to minimize $E^{\theta}(L) = \sum_{\theta \in \Theta} L(u, \theta) \cdot P(\theta)$

Example

$$\begin{bmatrix} & u_1 & u_2 & u_3 \\ \theta_1 & 1 & -1 & 2 \\ \theta_2 & -1 & 2 & -1 \\ \theta_3 & 0 & -2 & 1 \end{bmatrix}$$

Scenario 2: choose u_1, pay loss of 1 in the worst case.
Scenario 3: Let $P(\theta_1) = \frac{1}{5}$, $P(\theta_2) = \frac{1}{5}$, and $P(\theta_3) = \frac{3}{5}$. Then

$$E^{\theta}[L(u_1, \theta)] = \frac{1}{5}1 + \frac{1}{5}(-1) + \frac{3}{5}0 = 0$$

$$E^{\theta}[L(u_2, \theta)] = \frac{1}{5}(-1) + \frac{1}{5}(2) + \frac{3}{5}(-2) = -1$$

$$E^{\theta}[L(u_3, \theta)] = \frac{1}{5}(2) + \frac{1}{5}(-1) + \frac{3}{5}(1) = \frac{4}{5}$$

Choose u_2!!
What if $P(\theta_1) = \frac{1}{10}$, $P(\theta_2) = \frac{4}{5}$, and $P(\theta_3) = \frac{1}{10}$?
Then u_1 is chosen. Observe that the choice depends on $P(\theta)$. If this info is statistically valid, then better decisions are made. If not, then possibly worse decisions are made.

Example

$$\begin{bmatrix} & umbrella & no\ umbrella \\ rain & 1 & 5 \\ no\ rain & 1 & 0 \end{bmatrix}$$

Murphy's Law: always bring umbrella!!
Probabilistic: base decision on forecast.

Having a single observation

Let y be an observation (data, measurment, sensors).
Let Y be the observation space.
A decision now depends on y. Let γ denote a decision (strategy, plan). $\gamma : Y \rightarrow U$.

Nondeterministic: Assume there is some $F(Y) \subseteq \Theta$ which is known for every $y \in Y$. Choose γ that $m$inimizes

$$max_{\theta \in F(Y)} L(u, \theta)$$

for each $y \in Y$. Note that $u = \gamma(y)$.

Probabilistic: Assume $P(y|\theta)$ is known $\forall y \in Y$ and $\forall \theta \in \Theta$. Then

$$P(y|\theta) \cdot P(\theta) = P(\theta|y) \cdot P(y)$$

or equivalently

$$P(\theta|y) = \frac{P(y|\theta) \cdot P(\theta)}{P(y)}$$

$P(y|\theta)$ is given to us as well as $P(\theta)$. How do we find $P(y)$?

Soluition: Marginilization

$$P(Y) = \sum_{\theta \in \Theta} P(y|\theta) \cdot P(\theta)$$

Conditional Bayes Risk:

$$R(u|y) = \sum_{\theta \in \Theta} L(u, \theta) \cdot P(\theta|y)$$

Choose γ that minimizes

$$R(\gamma(y)|y) \ for \ every \ y \in Y$$

or

$$R(u|?) = \sum_{\theta \in \Theta} L(u, \theta) \cdot P(\theta|?)$$

k observations $y_1, y_2, \ldots, y_k$

$$R(u|y_1, \ldots, y_k) = \sum_{\theta \in \Theta} L(u, \theta) \cdot P(\theta|y_1, \ldots, y_k)$$

Assume $P(y_i|\theta)$ is known for each $i \in \{1, \ldots, k\}$. Assuming Conditional Independance then we have

$$P(\theta|y_1, \ldots, y_k) = \frac{P(y_1|\theta)P(y_2|\theta) \ldots P(y_k|\theta)}{P(y_1, y_2, \ldots, y_k)}$$

given θ the $y's$ are conditionally independant!!

An example of decision theory:
Classification
$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ set of classes.
y is the feature vector
Y is the feature space
$\Theta = U = \Omega$
Nature selects the class, ω, of an object.
Decision: choose the correct class.
Use:

$$L(u, \theta) = \left\{ \begin{array}{cc} 0 & if \ u = \theta \\ 1 & otherwise \end{array} \right\}$$

In the end we Minimize the probability of classification error!!