Lecture 9

Linear Programming

Ashish Dandekar

Prescriptive Analytics

Prescriptive analytics is the application of logic and mathematics to data to specify a preferred course of action. While all types of analytics ultimately support better decision making, prescriptive analytics outputs a decision rather than a report, statistic, probability or estimate of future outcomes.

- Gartner, Forecast Snapshot: Prescriptive Analytics

Prescriptive Analytics

Heuristics based

- Decisions that are purely devised by the domain experts.
- It may lead to a non-feasible solution.
- The formulation is not easily scalable.
- Insights are limited to the answers.

Optimisation based

- Rules are formalised using mathematical tools.
- It may lead to the best possible solution.
- The formulation is easily scalable.
- Insights can go beyond the best possible solution.

Linear Programming

Q1: ABC Woodcarving

- Demand for toy trains in unlimited but at most 40 soldiers are sold per week.
- A soldier is sold for $\$27$ and uses $\$10$ worth of raw materials. It also has a variable labour cost of $\$10$.
- A train is sold for $\$21$ and uses $\$9$ worth of raw materials. It also has a variable labour cost of $\$10$.
- A soldier requires 2 hours of finishing and 1 hour of carpentry.
- A train requires 1 hour of finishing and 1 hour of carpentry.
- Each week ABC can obtain all raw material but only 100 finishing hours and 40 carpentry hours.

Q1: ABC Woodcarving

Let's rearrange the same information in a better way.

	Soldier	Train	Constraints
Selling Cost	27	21	-
Raw Material Cost	10	9	-
Overhead Cost	14	10	-
Carpentry Labour	1	1	80
Finishing Labour	2	1	100
Demand	≤ 40	-	-

We want to find the number of soldiers and trians to be manufactured such that the profit is maximised.

Optimisation Problem

Decision variable

Decision variables that can be controlled by the decision makers.

Objective function

Objective function is a mathematical function of the decision variables that converts a solution into a numerical evaluations.

Question 1.

- x_1 be the number of soldiers manufactured every week.
- x_2 be the number of trains manufactured every week.

$$
\begin{aligned}
&\mathit{profit}(x_1, x_2) \\
&= (27 - 10 - 14)x_1 + (21 - 9 - 10)x_2 \\
&= 3x_1 + 2x_2
\end{aligned}
$$

Thus,

$$\max_{x_1, x_2 \in Z} 3x_1 + 2x_2$$

Constraints

Constraints

A set of functional equalities or inequalities that quantify the restrictions on the values of the decision variables.

	Soldier	Train	Constraints
Selling Cost	27	21	-
Raw Material Cost	10	9	-
Overhead Cost	14	10	-
Carpentry Labour	1	1	80
Finishing Labour	2	1	100
Demand	≤ 40	-	-

$$
\begin{aligned}
2x_1 + x_2 &\leq 100 \\
x_1 + x_2 &\leq 80 \\
x_1 &\leq 40 \\
x_1, x_2 &\geq 0
\end{aligned}
$$

Linear Programming

Linear program is an optimisation problem where

- The objective function is a *linear* function of decision variables.
- The values of the decision variables must satisfy a set of *linear constraints*.
- A *sign restriction* is associated with every decision variable.

$$\max_{x_1, x_2 \in Z} 3x_1 + 2x_2$$

$$2x_1 + x_2 \leq 100$$

$$x_1 + x_2 \leq 80$$

$$x_1 \leq 40$$

$$x_1, x_2 \geq 0$$

Assumptions

Proportionality

The contribution of individual decision variables in the objective function is (only) proportional to their value in it.

Additivity

The contribution to the objective function for any decision variable is independent of the values of the other decision variables.

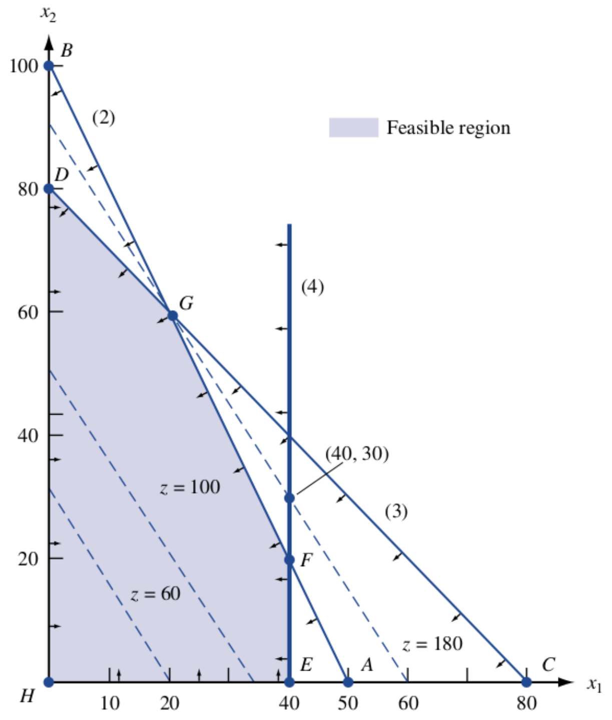
Divisibility

The decision variables can take any real values within the specified range.

Certainty

Ever parameter value in the model is know with the certainty.

Geometry of LP



Feasible Region

It is the set of all points that satisfy every constraint in the linear program.

Optimal Solution

It is a point in the feasible region with the largest (or smallest) value for the optimisation function.

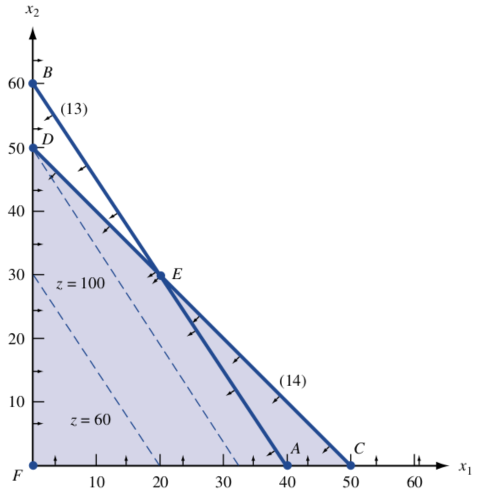
Optimal solution always lies at one of the corners of the polyhedron defined by the constraints!

Binding Constraint

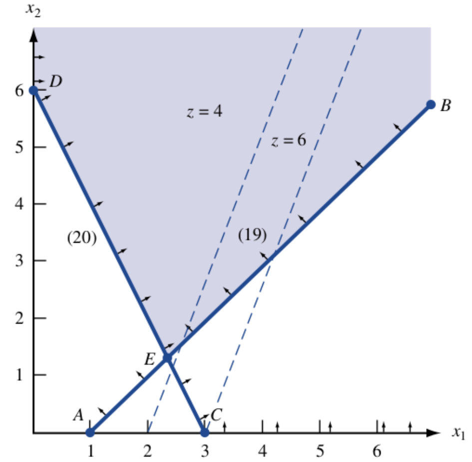
The constraint is said to be binding if the optimal solution on the line defined by the constraint.

Does (unique) solution always exist?

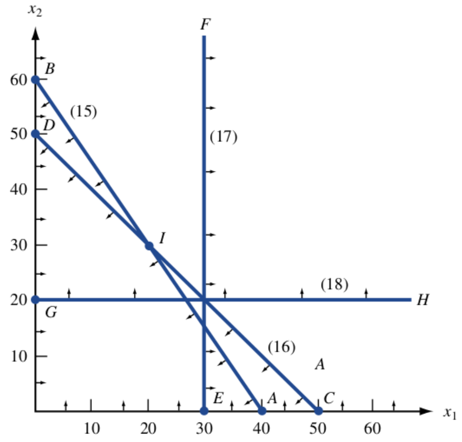
Multiple Optimal Solutions



Unbounded LP



Infeasible LP



Q2: Job Scheduling

A post office requires different number of employees on different days of the week.

Constraint from the Union

Each employee must get two off days after working for five consecutive days.

Objective

Minimise the number of employees.

Day	#Employees
Monday	17
Tuesday	13
Wednesday	15
Thursday	19
Friday	14
Saturday	16
Sunday	11

Q2: Job Scheduling

Let, x_i be the number of employees that start work on day i.

Who works on Monday?

Everybody except those who started on Tuesday and Wednesday.

$$x_1 + x_4 + x_5 + x_6 + x_7 \geq 17$$

Day	#Employees
Monday	17
Tuesday	13
Wednesday	15
Thursday	19
Friday	14
Saturday	16
Sunday	11

Q2: Job Scheduling

Linear Program

$$\min x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7$$

$$x_1 + x_4 + x_5 + x_6 + x_7 \geq 17$$
$$x_2 + x_5 + x_6 + x_7 + x_1 \geq 13$$
$$x_3 + x_6 + x_7 + x_1 + x_2 \geq 15$$
$$x_4 + x_7 + x_1 + x_2 + x_3 \geq 19$$
$$x_5 + x_1 + x_2 + x_3 + x_4 \geq 14$$
$$x_6 + x_2 + x_3 + x_4 + x_5 \geq 16$$
$$x_7 + x_3 + x_4 + x_5 + x_6 \geq 11$$
$$x_1, x_2, x_3, x_4, x_5, x_6, x_7 \geq 0$$

Day	#Employees
Monday	17
Tuesday	13
Wednesday	15
Thursday	19
Friday	14
Saturday	16
Sunday	11

Q2: Job Scheduling

The solver provides the optimal solution:

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = (\frac{4}{3}, \frac{10}{3}, 2, \frac{22}{3}, 0, \frac{10}{3}, 5)$$
$$\approx (2, 4, 2, 8, 0, 4, 5)$$

An optimal solution with value 23 exists!

Integer Linear Program (ILP)

ILP violates the divisibility assumption of the linear programs. Integer programming is NP-Complete. Relaxing ILP (integer) to LP (real) is an approximation which may not lead to optimal solution.

Q3: Perfume Production Process

Raw Material

- Purchased at $\$3$ per pound.
- Requires 1 hour of lab processing.
- Every poind yields 3oz of *regular brute* and 4oz of *regular chanelle*.
- Every year the company gets 6000 lab hours.
- Every year the company purchases at most 6000 pounds of raw materials.

Selling cost

Blend	Price per oz
Regular brute	$\$7$
Regular chanelle	$\$6$
Luxury brute	$\$18$
Luxury chanelle	$\$14$

Q3: Perfume Production Process

Manufacturing luxury brute

- An ounce of regular brute yields an ounce of luxury brute.
- Requires 3 additional lab hours.
- Takes $\$4$ of processing cost.

Manufacturing luxury chanelle

- An ounce of regular chanelle yields an ounce of luxury chanelle.
- Requires 2 additional lab hours.
- Takes $\$4$ of processing cost.

Objective is to maximise profit!

Q3: Perfume Production Process

Decision Variable

x_1 = ounces of regular brute sold annually

x_2 = ounces of luxury brute sold annually

x_3 = ounces of regular chanelle sold annually

x_4 = ounces of luxury chanelle sold annually

x_5 = pounds of raw material purchased annually

Q3: Perfume Production Process

Blend	Price per oz
Regular brute	$7
Regular chanelle	$6
Luxury brute	$18
Luxury chanelle	$14

Objective Function

$$\max 7x_1 + (18-4)x_2 + 6x_3 + (14-4)x_4 - 3x_5$$

Each luxury blend bears $4 *of processing cost.*

Q3: Perfume Production Process

Constraints

$$\begin{aligned} x_5 &\leq 4000 \quad \text{(Raw material)} \\ x_5 + 3x_2 + 2x_4 &\leq 6000 \quad \text{(Laboratory hours)} \end{aligned}$$

Mixing constraints

- A pound of raw material yields $3oz$ of brute

$$x_1 +_x 2 = 3x_5$$

- A pound of raw material yields $5oz$ of chanelle

$$x_3 +_x 4 = 5x_5$$

Q3: Perfume Production Process

Putting it all together!

$$\max 7x_1 + (18-4)x_2 + 6x_3 + (14-4)x_4 - 3x_5$$

subject to

$$\begin{aligned} x_5 &\leq 4000 \\ x_5 + 3x_2 + 2x_4 &\leq 6000 \\ x_1 +_x 2 - 3x_5 &= 0 \\ x_3 +_x 4 - 5x_5 &= 0 \\ x_1, x_2, x_3, x_4, x_5 &\geq 0 \end{aligned}$$

Q4: Portfolio management

The cashflow associated with the investment of $\$1$ in the investment schemes is given as follows:

Schemes	Cash flow at time			
	0	**1**	**2**	**3**
A	-1	0.5	1	0
B	0	-1	0.5	1
C	-1	1.2	0	0
D	-1	0	0	1.9
E	0	0	-1	1.5

- At time $t = 0$, $\$100k$ are available for the investment.
- Finco requires at most $\$75k$ to be invested in a single scheme.
- Finco earns 8% interest by keeping univested cash in the market funds.

We want ot maximise the cashflow at the end of third year!

Q4: Portfolio management

Decision Variables

A, B, C, D, E be the amount in dollars invested in the scheme A, B, C, D, E respectively at time $t = 0$.

S_t be the amount invested in market funds at time t

Objective Function

$$\max B + 1.9D + 1.5E + 1.08S_2$$

Schemes	Cash flow at time			
	0	**1**	**2**	**3**
A	-1	0.5	1	0
B	0	-1	0.5	1
C	-1	1.2	0	0
D	-1	0	0	1.9
E	0	0	-1	1.5

Q4: Portfolio management

Constraints

$$A + C + D + S_0 = 10000$$
$$0.5A + 1.2C + 1.08S_0 = (B + S_1)$$
$$A + 0.5B + 1.08S_1 = E + S_2$$
$$A, B, C, D, E \leq 75000$$
$$A, B, C, D, E \geq 0$$

Schemes	Cash flow at time			
	0	**1**	**2**	**3**
A	-1	0.5	1	0
B	0	-1	0.5	1
C	-1	1.2	0	0
D	-1	0	0	1.9
E	0	0	-1	1.5

Sensitivity Analysis

Sensitivity analysis focuses on how change in LP's parameters changes the optimal solution.

It can be studies in two ways:

- Change in the availability of the resources (the constraints)
- Change in the per unit profit or cost (coefficients of the objective function)

Sensitivity Analysis

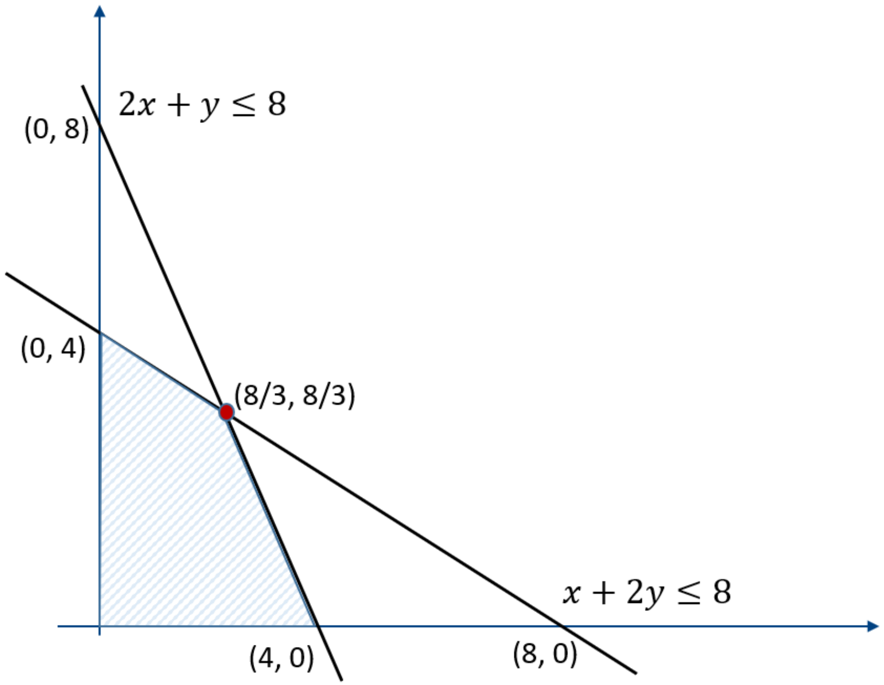
	Product 1	Product 2	Availability
Machine 1	2hrs	1hr	8hrs
Machine 2	1hr	2hrs	8hrs
Profit/Unit	300	200	

Let x and y denote the number of Product 1 and Product 2 manufactured respectively.

$$\max_{x,y} 300x + 200y$$

$$\begin{aligned} s.t.\;\; 2x + y &\leq 8 \\ x + 2y &\leq 8 \\ x, y &\geq 0 \end{aligned}$$

Sensitivity Analysis

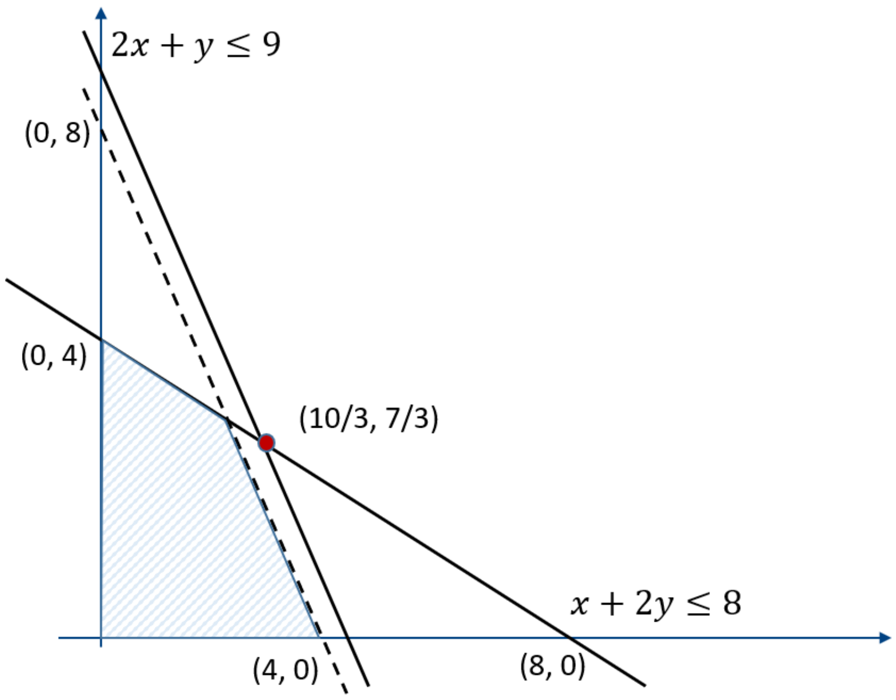


$$\max_{x,y} 300x + 200y$$

$$\begin{aligned} s.t.\ \ 2x + y &\leq 8 \\ x + 2y &\leq 8 \\ x, y &\geq 0 \end{aligned}$$

Optimal answer is: $4000/3$

Sensitivity Analysis



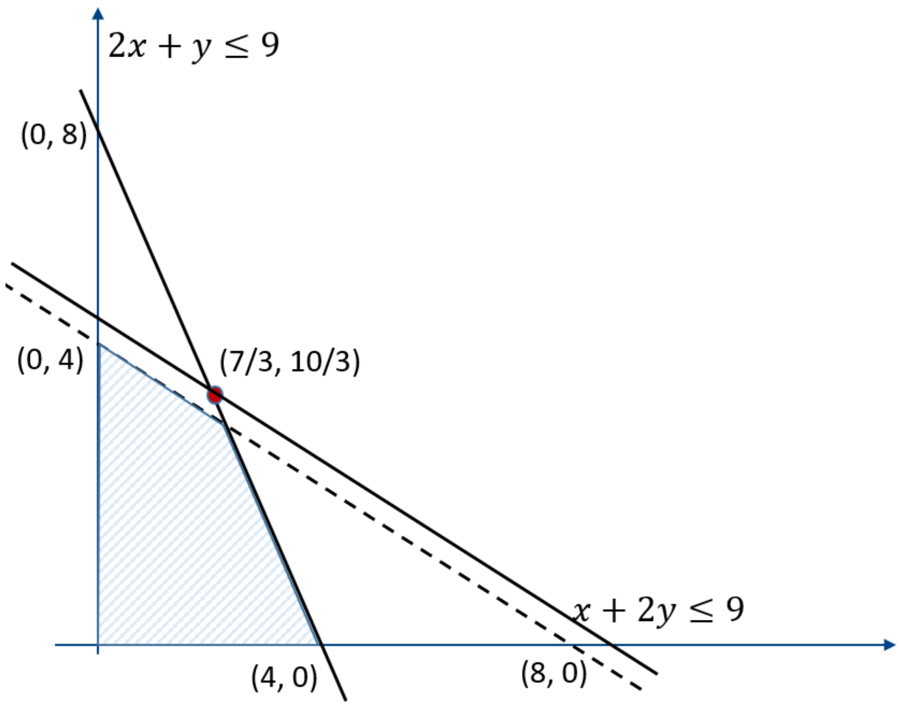
What if Machine 1 is available for 9 hours?

$$\max_{x,y} 300x + 200y$$

$$\begin{aligned} s.t.\ 2x + y &\leq 9 \\ x + 2y &\leq 8 \\ x, y &\geq 0 \end{aligned}$$

Optimal answer is: $4400/3$

Sensitivity Analysis



What if Machine 2 is available for 9 hours?

$$\max_{x,y} 300x + 200y$$

$$\begin{aligned} s.t.\ \ 2x + y &\leq 8 \\ x + 2y &\leq 9 \\ x, y &\geq 0 \end{aligned}$$

Optimal answer is: $4100/3$

Sensitivity Analysis

Shadow Price

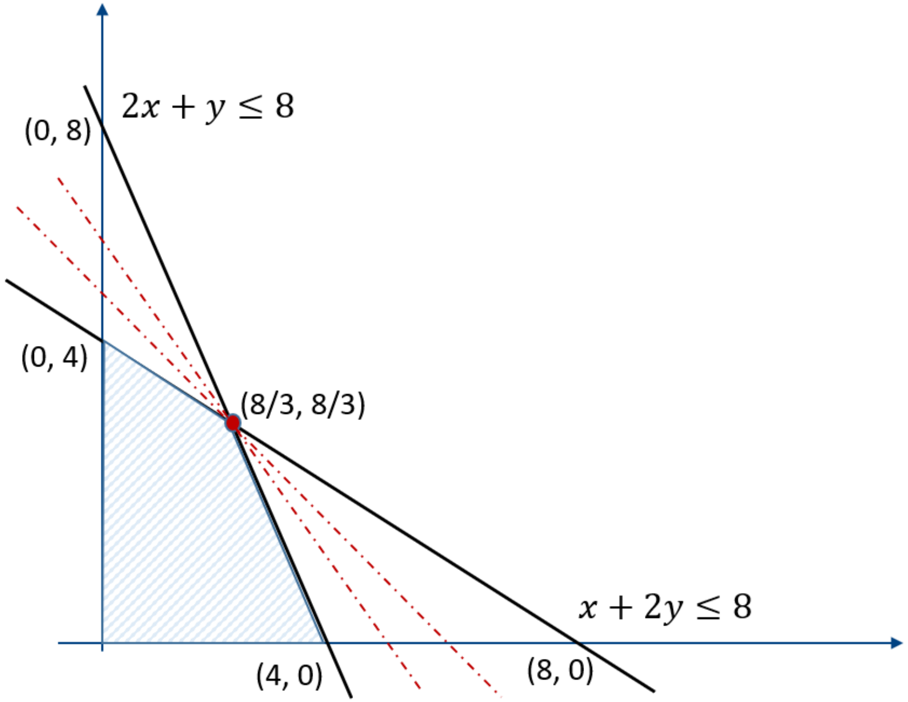
Shadow price for the i^{th} constraint is defined as the change in the value of the objective function per unit change in the constraint.

$$\text{Shadow price for Machine 1} = \frac{4400}{3} - \frac{4000}{3} \approx 133$$
$$\text{Shadow price for Machine 2} = \frac{4100}{3} - \frac{4000}{3} \approx 33$$

Prescriptive Analytics

- Which machine should be given priority?
- What if the operating cost of a machine changes by \$50?

Optimality Range



Optimality range gives us the range of coefficients of objective function such that the optimal solution does not change.

Slope of the line

Slope of line $ax + by$ is is $-a/b$

$$\frac{1}{2} \leq \frac{p_1}{p_2} \leq 2$$

Optimality Range

- Suppose that we change the per unit profit of the product to $\$350$ and $\$250$. Will it change the optimal solution?

$$\frac{p_1}{p_2} = \frac{7}{5} \leq 2$$

- Suppose that we change the profit per unit of the Product 2 to $\$200$. What range of the profit per unit of the Product 1 that will not alter the optimal solution?

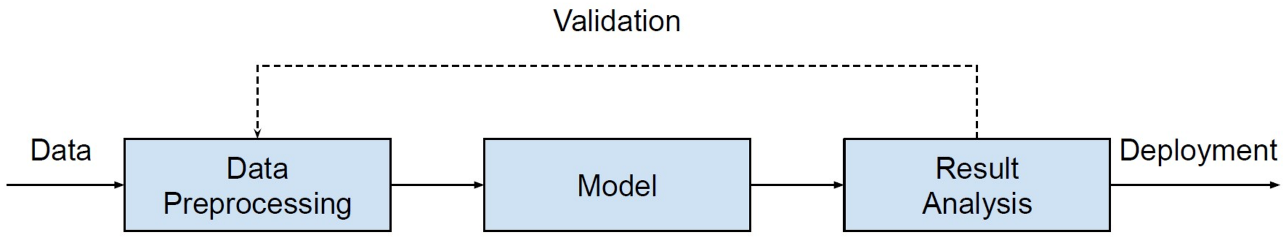
$$\frac{1}{2} \leq \frac{p_1}{200} \leq 2 \quad \Rightarrow \quad 100 \leq p_1 \leq 400$$

Summary

- Prescriptive analytics as optimisation
- Linear Programming

Looking back...

Data Analytics Pipeline



Data preprocessing

- Data cleaning
- Data transformation
- Feature selection

Modeling

- Regression
- Classification
- Clustering

Result analysis

- Qualitative
- Quantitative
- Comparative

Week 1 to Week 3

Statistics

- Mesures of central tendancies
- Confidence intervals
- Hypothesis testing

Data Preprocessing

- Data cleaning
- Data transformation
- Dimensionality Reductions

Week 4 to Week 8

Recipe Machine Learning

- Hypothesis set
- Minimisation of loss over the training set
- Bayesian view (everything is a random variable)
- Maximisation of likelihood (or posterior)

Regression Analysis

- Linearity, Normality, Homoscadasticity and Independence
- Regularisation
- Difference in Differences Analysiss

Classification Analysis

- Linear models and tree models
- Confusion matrix with assoiated metrics
- Ensemble models

Week 9

Linear programming

- Introduction to Prescriptive Analytics
- Formalising the optimisation problem

Assessments

Assignment 1

- Targeted at the descriptive analytics
- Focus is on qualitative analysis

Assignment 2

- Targetd at predictive analytics
- Focus is on quantitative analysis

Project

- Taste of real-world data
- Balance of qualitative and quantitative analysis

Midterm

- Designed to assess the coding skills

Final exam

- Designed to assess the analytical skills
- Designed to assess the theoretical fundamentals of data analytics

What's next?

Core ML

- Research oriented work. Designing new ML models

Data Scientist

- Putting ML and statistics in the context of Data

Data Engineering

- Preparing data pipelines and working at scale

ML Engineering

- Training and deployment of models at scale

Business Analyst