Business Analytics

Session 10b. Integer Programming

Renyu (Philip) Zhang

New York University Shanghai

Spring 2019

Integer Solutions to LP

Rounding LP Solutions

- When is rounding a non-integer solution to an LP acceptable?
- If variables take on large values, rounding usually does not have big impact on feasibility or optimality.
  - This is the case for Google AdWords.
- Rounding is a big problem if the optimal LP solution values are small.
  - Especially when the decisions are 1/0, Yes/No.
- A new approach: Interger programming.

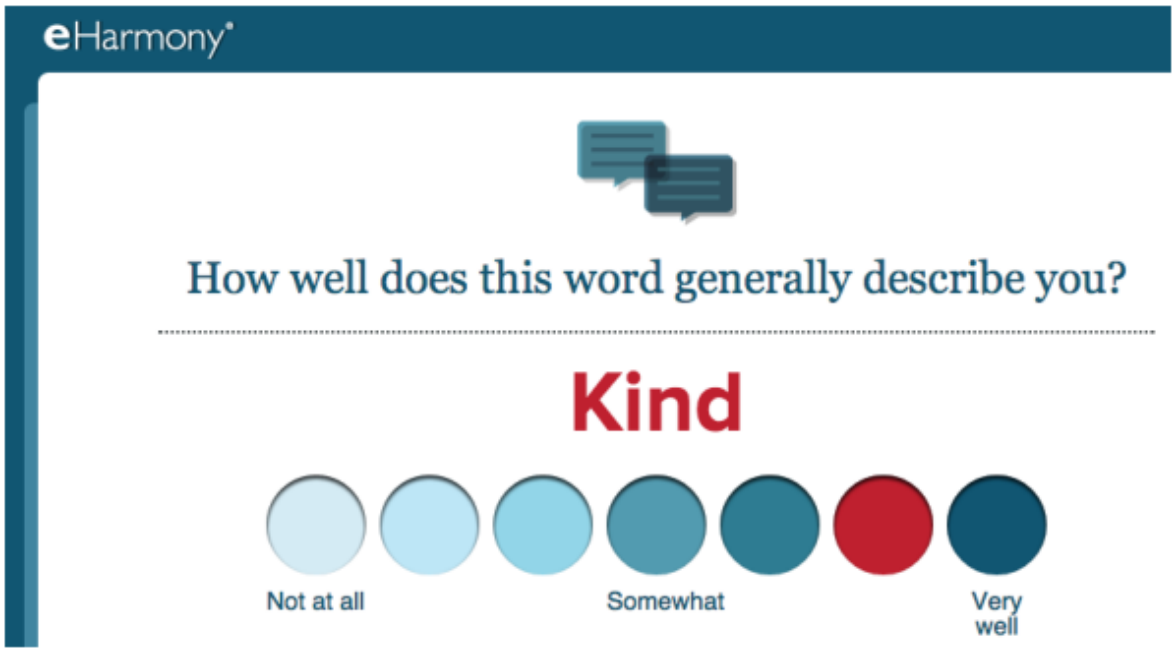
eHarmony

Business Model of eHarmony

- Online dating website focusing on long term relationships.
- Computes a compatibility score between two people, and uses optimization models to determine the best matches.
- Successful business
  - 4% of US marriages are a result of eHarmony
  - 14% of US online dating market
  - Over $1 billion revenue
  - Divorce rate: 3.86%

Compatibility Scores

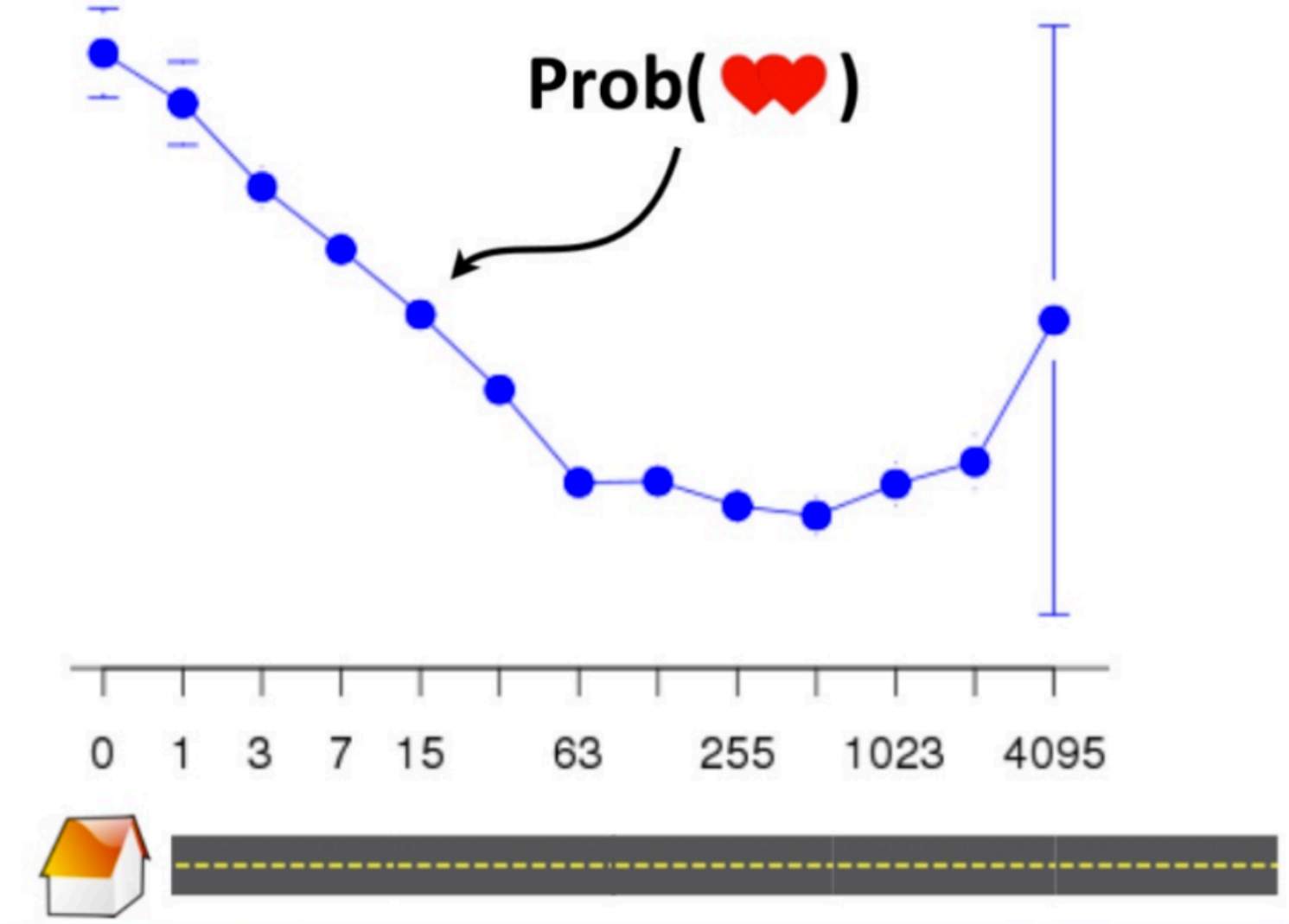
- Questionnaire of 436 questions on 29 different dimensions of personality.
- Computes a compatibility score between two people, and uses optimization models to determine the best matches.
- Successful matches must meet 25/29 compatibility areas.



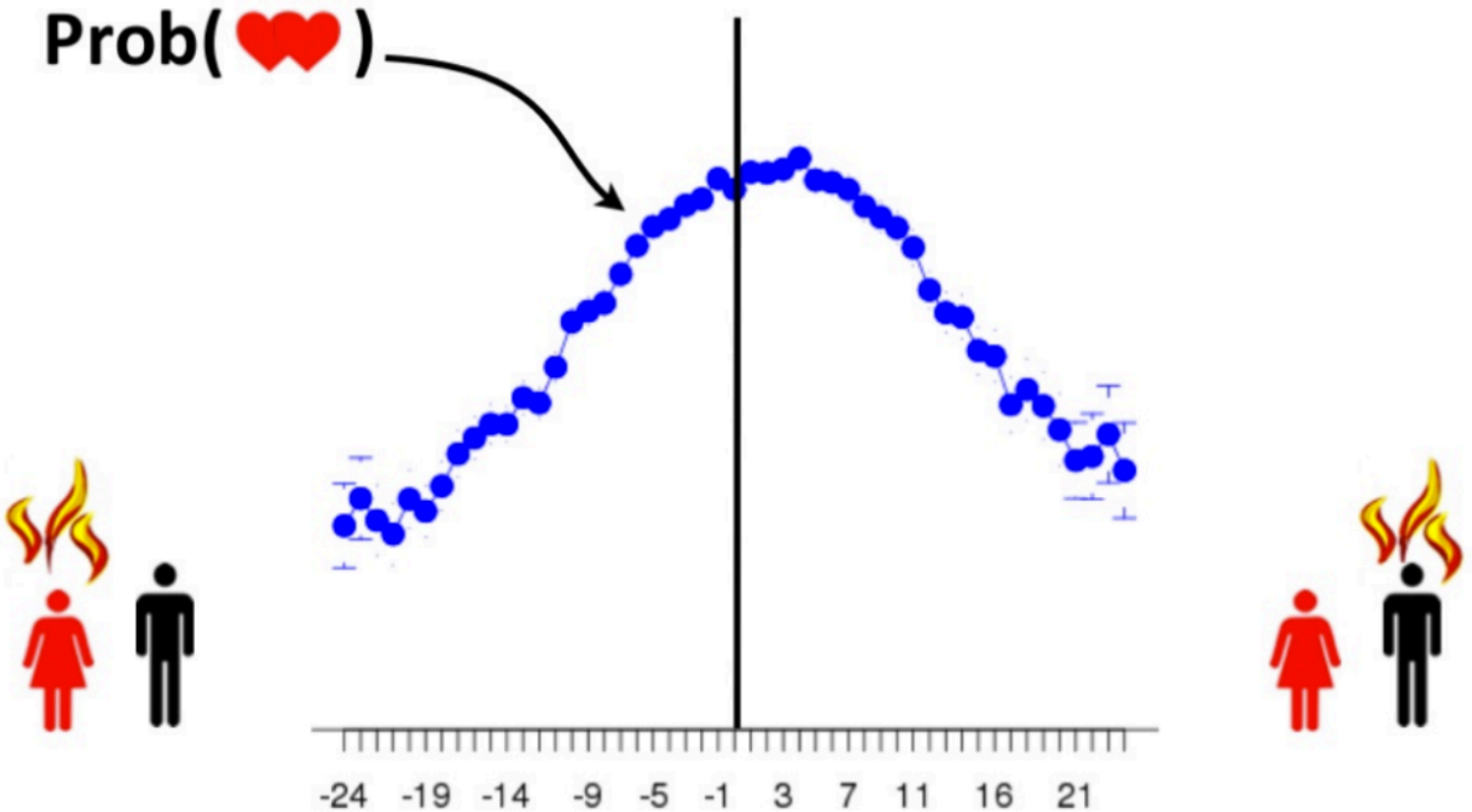
Compatibility Scores

- Clustering: Compatibility scores based on similarity between users' answers to the questionnaire.
- Trees: Compatibility between users bears a nonlinear structure.
- Text analytics: Analyze the text of users' profiles.

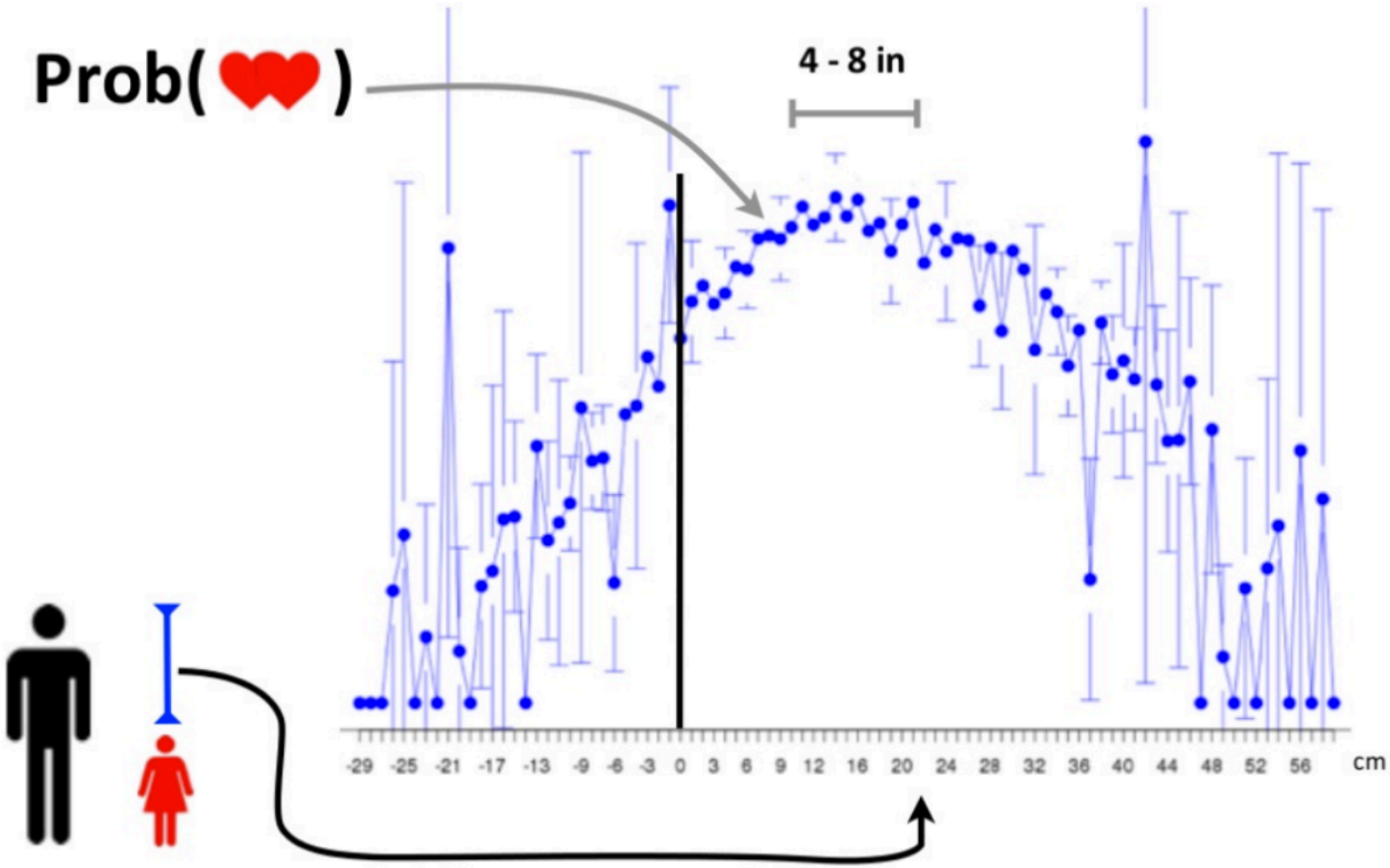
Distance



Attractiveness of Appearances



Height Difference

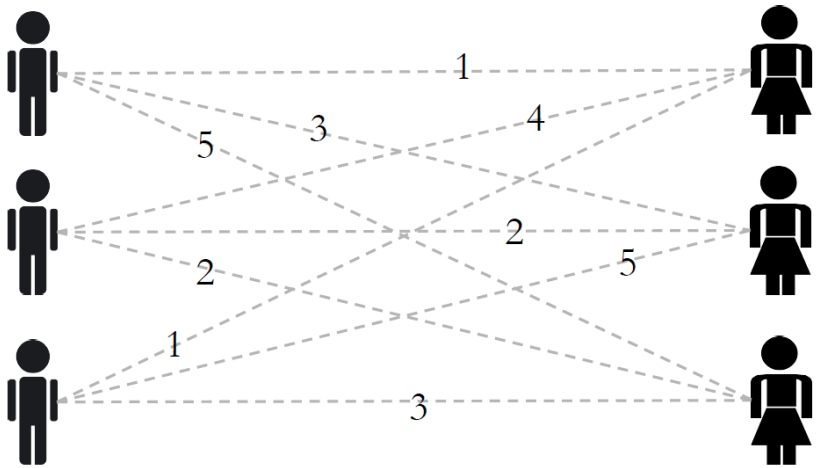


eHarmony's Matching Problem

Matching Optimization

- Goal of eHarmony: Determine who should be matched to whom given the compatibility scores

- Toy example: 3 men and 3 women; compatibility score between 1 and 5 for all pairs.



	W_1	W_2	W_3
M_1	1	3	5
M_2	4	2	2
M_3	1	5	3

Table 1: Compatibility Scores

Matching Optimization

- Key question: How to match pairs together to maximize the total compatibility score?
- Each woman should be matched with exactly one man, and vice versa.

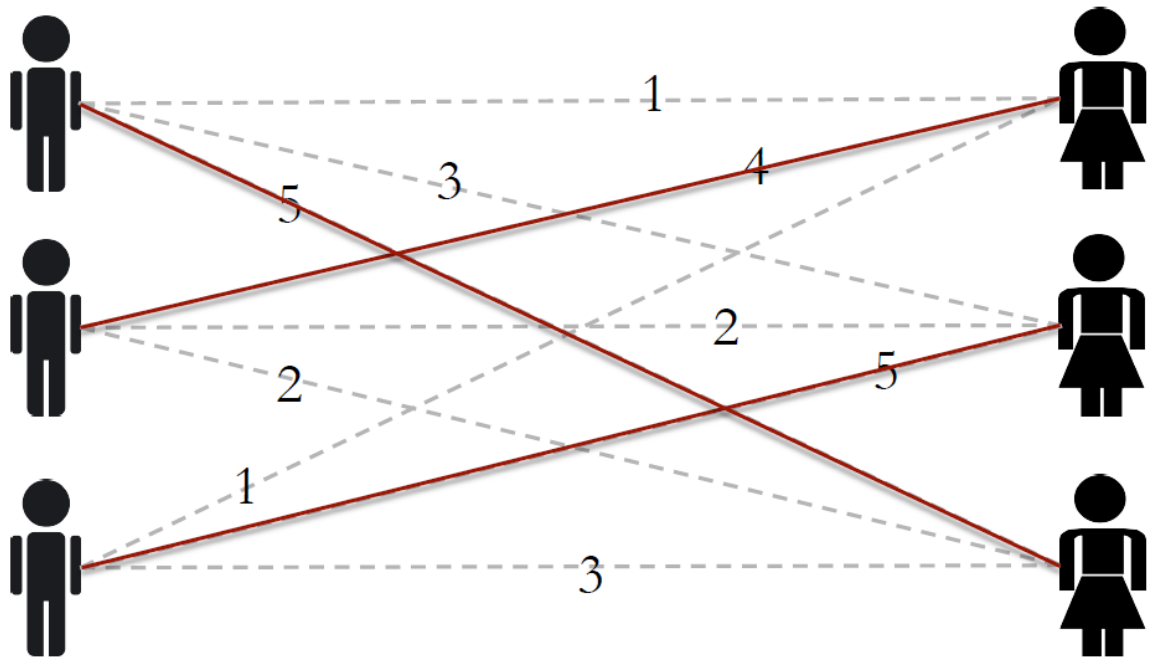


Figure 1: A Feasible Matching

Greedy Strategy May not Be Optimal

	W_1	W_2
M_1	5	3
M_2	4	1

Table 2: Compatibility Scores

- Greedy matching: $M_1 - W_1$, $M_2 - W_2$.
  - Total compatibility score: $5 + 1 = 6$
- Optimal matching: $M_1 - W_2$, $M_2 - W_1$.
  - Total compatibility score: $4 + 3 = 7$

Integer Programming: Decision Variables and Objective Function

	W_1	W_2	W_3
M_1	x_{11}	x_{12}	x_{13}
M_2	x_{21}	x_{22}	x_{33}
M_3	x_{31}	x_{32}	x_{33}

- $x_{ij} = 1$ if we match M_i and W_j together; otherwise $x_{ij=0}$.
  - For the case in Figure 1, $x_{13} = x_{21} = x_{32} = 1$, and the rest of $x_{ij} = 0$.
- This is called an integer programming model as the decision variables x_{ij}'s only take integer (0 or 1) values.
- Objective function:

$$x_{11} + 3x_{12} + 5x_{13} + 4x_{21} + 2x_{22} + 2x_{23} + x_{31} + 5x_{32} + 3x_{13}$$

Integer Programming: Constraints

- Binary constraints: $x_{ij} \in \{0, 1\}$ $(1 \leq i, j \leq 3)$.
- Each man is matched with one woman:
  - $x_{11} + x_{12} + x_{13} = 1$
  - $x_{21} + x_{22} + x_{23} = 1$
  - $x_{31} + x_{32} + x_{33} = 1$
- Each woman is matched with one man:
  - $x_{11} + x_{21} + x_{31} = 1$
  - $x_{12} + x_{22} + x_{32} = 1$
  - $x_{13} + x_{23} + x_{33} = 1$

Integer Programming (IP) for eHarmony Matching

$$\max\ x_{11} + 3x_{12} + 5x_{13} + 4x_{21} + 2x_{22} + 2x_{23} + x_{31} + 5x_{32} + 3x_{13}$$

Subject to

$$\begin{aligned}
x_{11} + x_{12} + x_{13} &= 1 \\
x_{21} + x_{22} + x_{23} &= 1 \\
x_{31} + x_{32} + x_{33} &= 1 \\
x_{11} + x_{21} + x_{31} &= 1 \\
x_{12} + x_{22} + x_{32} &= 1 \\
x_{13} + x_{23} + x_{33} &= 1 \\
x_{ij} \in \{1, 0\},\ &\text{for } 1 \leq i, j \leq 3
\end{aligned}$$

IP for eHarmony Matching: Larger Scale

- N men and N women on the platform.
- The compatibility score of M_i to W_j is s_{ij}.
- $x_{ij} = 1$ if we match M_i to W_j; otherwise $x_{ij} = 0$.

$$\max \sum_{i=1}^{N} \sum_{j=1}^{N} s_{ij} x_{ij}$$

Subject to

$$\sum_{j=1}^{N} x_{ij} = 1 \text{ for all } 1 \leq i \leq N \text{ (Each man is matched with one woman)}$$

$$\sum_{i=1}^{N} x_{ij} = 1 \text{ for all } 1 \leq j \leq N \text{ (Each woman is matched with one man)}$$

$$x_{ij} \in \{1, 0\}, \text{ for } 1 \leq i, j \leq N$$

Solving the Integer Program in Python

- Use the package "cvxopt" in Python to solve this integer program.
- Demonstration in Python.
- Optimal Matching Strategy:

	W_1	W_2	W_3
M_1	$x^*_{11} = 0$	$x^*_{12} = 0$	$x^*_{13} = 1$
M_2	$x^*_{21} = 1$	$x^*_{22} = 0$	$x^*_{33} = 0$
M_3	$x^*_{31} = 0$	$x^*_{32} = 1$	$x^*_{33} = 0$

- Optimal Compatibility Score=14
- The IP model for eHarmony's matching optimization problem is also called the assignment model.
  - Even if we allow $x_{ij} \in [0, 1]$, the optimal solution would still be binary, i.e., $x^*_{ij} \in \{0, 1\}$.

Another Example of Integer Programming

Capital Budgeting

- The Tatham Company is considering 7 investment options, the cash requirements and net present value (NPV) of which are:

Project	Cash Required	NPV
1	$5,000	$16,000
2	$2,500	$8,000
3	$3,500	$10,000
4	$6,000	$19,500
5	$7,000	$22,000
6	$4,500	$12,000
7	$3,000	$7,500

- The cash available for investment is $15,000.
- Goal: Maximize the total NPV.
- Formulate the capital budgeting problem as an integer program.

Modeling Constraints in Integer Programming

- If project 4 is undertaken, then project 5 must be taken: $x_4 \leq x_5$.
- If project 4 and project 5 must be undertaken together: $x_4 = x_5$.
- If project 3 is undertaken, project 4 must be rejected: $x_3 + x_4 \leq 1$.
- At most 3 of projects 1 through 5 can be undertaken:
  $x_1 + x_2 + x_3 + x_4 + x_5 \leq 3$.
- Exactly 1 of the first 3 projects must be undertaken:
  $x_1 + x_2 + x_3 = 1$.

Homework

- Submit your choice of final topic to me if you haven't done so.
- Read *Analytics Edge*, Chapter 12, Chapter 22.6-22.7.
- Finish Homework 9 and 10 (NO need to submit it).
- Read the required reading for Session 11.