14. **(5 points)** When applying the recommendation algorithms, what do we achieve by normalization of each row of the utility matrix (subtract the mean from rating values)?

What we achieve when the act of normalization occurs on each row is "normed" so we're seeking relations (on same scale). {similarity between data}

15. **(5 points)** Please list two drawbacks for both content-based filtering and collaborative filtering approaches.

Content Based filtering	Collaborative
(1) The model relies heavily on existing interests from the user. It cannot expand them.	(1) Cannot handle new introduced items not seen in training.
(2) Determining what Others users think of the item is not taken into account.	(2) "Cold Start" the algorithm relies on users in the system to cross reference.

- **Machine Learning (25 points)**

16. **(8 points)** Suppose we have a binary classification data with classes $Y \in \{+1, -1\}$ and d features with each feature $f_i \in \{+1, -1\}$. To improve the performance of the classifier, Jana decided to duplicate each feature. Hence, each training example now has $2d$ features with $f_{d+i} = f_i$ for $i = 1, 2, \cdots, d$. This question is about comparing the training problem with *original* feature set and *double* feature set. Assume that there are same number of training examples for both positive and negative class, and in case of ties, you will chose positive class.

For a Perceptron classifier, select all that apply.

$Y = \{1, -1\}$

a) Test accuracy with original feature set could be higher

b) Test accuracy with double feature set could be higher

d features

(c) Test accuracy will be same with both original and double feature set

Please write one sentence justification

The testing accuracy should not be changed simply by having a duplicate feature (on each). Especially since the rules are fixed i.e. ties = positive

Consider the following Perceptron, for which the inputs are the always "1" feature and two binary features $x_1 \in \{0, 1\}$ and $x_2 \in \{0, 1\}$. The output label $y \in \{0, 1\}$. Suppose w_0, w_1, w_2 stands for weights of the three features. The classification decision is made as follows: $y = 1$ if $(w_0 + w_1 \cdot x_1 + w_2 \cdot x_2) > 0$. Otherwise, $y = 0$.

17. **(5 points)** Which of the following choices for the weight vector (w_0, w_1, w_2) can classify y as $y = (x_1 \text{ XOR } x_2)$? XOR is the logical exclusive OR operation, which equals to ZERO when x_1 equals to x_2, and equals to ONE when x_1 is different from x_2.

$w_0 - w_2$ weights

x_1	x_2	y (XOR)
0	1	1
1	0	1
1	1	0

a) (1, 1, 0)

b) (-2, 1, 1.5)

c) Any weights that satisfy $(-w_1 - w_2) < w_0 < min(0; -w_1; -w_2)$

(d) No weights can compute the XOR logical relation

No weights seem able to satisfy the $y = (x_1 \text{ XOR } x_2)$ when used a perceptron model

$w_1 x_1 \ldots$