# 01 Learning from a large number of feature combinations

## CFM: low-rank regression with global optimal guarantees

### Abstract

Convex Factorization Machines (CFM) are a high-accuracy regression model that can handle a large number of feature combinations. CFM is general-purpose and can be applied to a wide range of tasks: e.g., house price prediction, recommender systems and genome analysis.

The proposed method can handle a large number of feature combinations by using a low-rank constraint. Moreover, it is guaranteed to obtain a global optimum

In future work, to further improve predictive accuracy, we plan to support higher-order feature combinations. Besides recommender systems, applications include predicting combinations of genes that are responsible for diseases, which would help find effective cures.

## Convex Factorization Machines (CFM): high-accuracy regression model that can find important feature combinations

- General-purpose (large-number of applications)
- Parameter learning is easy (insensitive to initialization)
- Can handle a large number of feature combinations (scales to high-dimensional data)
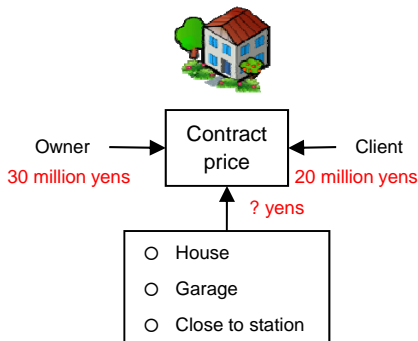- Can estimate the weight of unobserved feature combinations (can discover new knowledge)

$$\hat{y} = w^{\mathrm{T}} x + x^{\mathrm{T}} W x$$

Target variable — 1st order weights — Features — 2nd order weights — Large matrix!

CFM reduce the size of W using a low-rank constraint



$$W = \lambda_l \, p_l \, p_l^{\mathrm{T}} + \ldots + \lambda_k \, p_k \, p_k^{\mathrm{T}}$$

W is estimated in order of increasing eigenvalues

| Application 1 | Application 2 | Application 3 |
|---|---|---|
| Predict house price | Recommend books to users | Predict corn yield from DNA |



Owner 30 million yens → Contract price ← Client 20 million yens

? yens

- House
- Garage
- Close to station

Discover combinations of factors which impact house price

|  |  |  |
|---|---|---|
| Alice | ☆ | ☆☆☆ |
| Bob | ☆☆☆ | ? |
| Charlie | ? | ☆☆ |

Predict ratings from user and book interactions



Genes correlated with corn yield

8 tons / hectare    12 tons / hectare
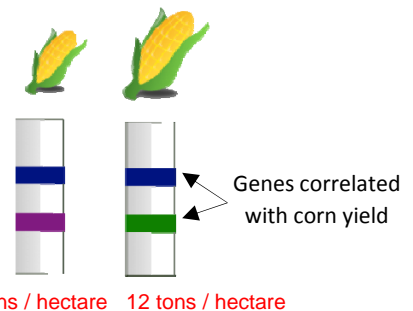
Predict corn yield from gene interactions!

【Reference】

[1] M. Blondel, A. Fujino, N. Ueda, "Convex factorization machines," in Proc. *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2015.

【Contact】

**Mathieu Blondel** Learning and Intelligent Systems Research group

E-mail：mathieu.blondel(at)lab.ntt.co.jp