Linear Quiz Blending

u++ (@upura0)

May 23, 2019

コンペ解法の論文

The BigChaos Solution to the Netflix Grand Prize

Andreas Töscher and Michael Jahrer

commendo research & consulting
Neuer Weg 23, A-8580 Köflach, Austria
{*andreas.toescher,michael.jahrer*}@*commendo.at*

Robert M. Bell*

AT&T Labs - Research
Florham Park, NJ

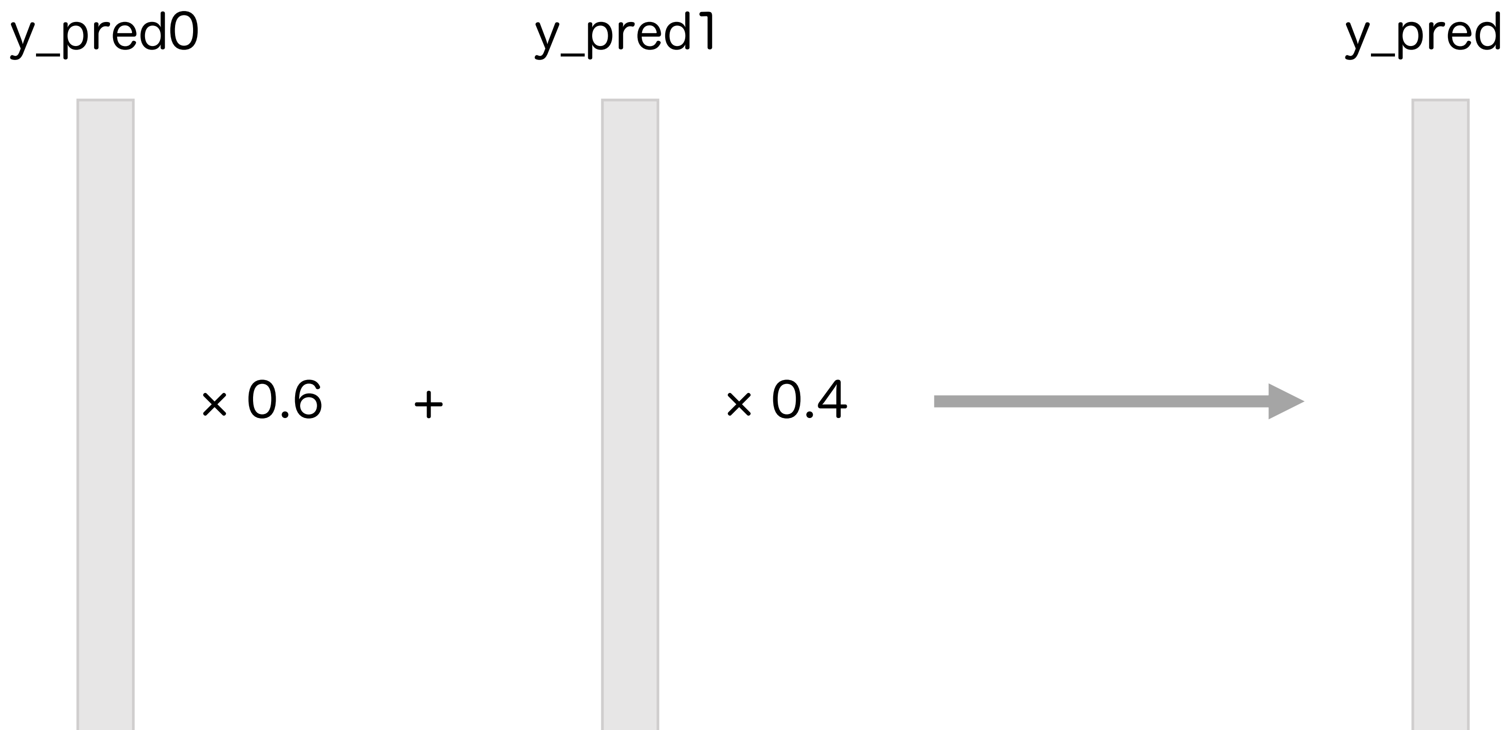
September 5, 2009

https://www.netflixprize.com/assets/GrandPrize2009_BPC_BigChaos.pdf

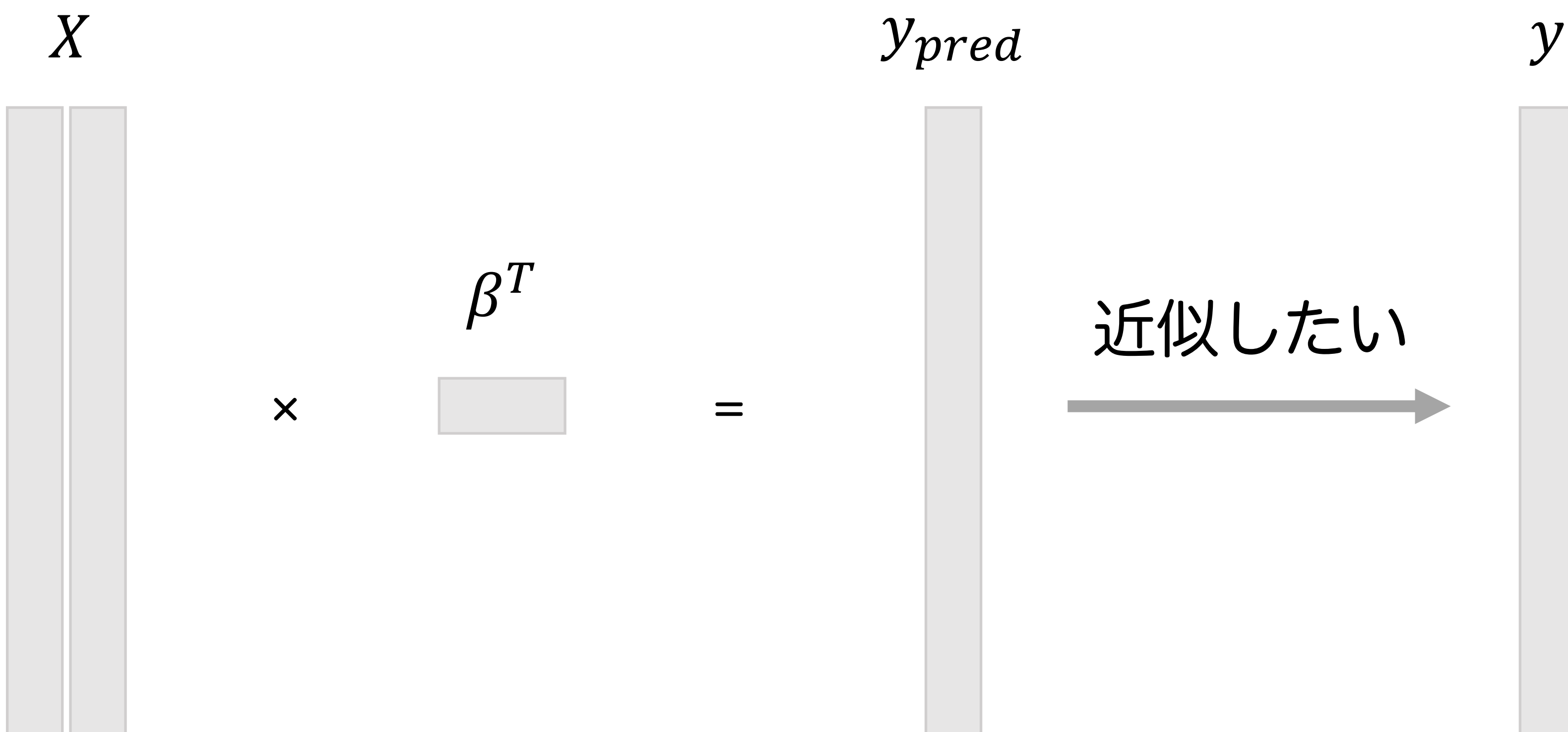
Linear Quiz Blending の概要

- public LBの情報を用いて、優れたensembleの重みを近似計算する
- 評価指標が最小二乗誤差の場合に利用できる

良い予測値の混ぜ方？



定式化 (良いβを知りたい)



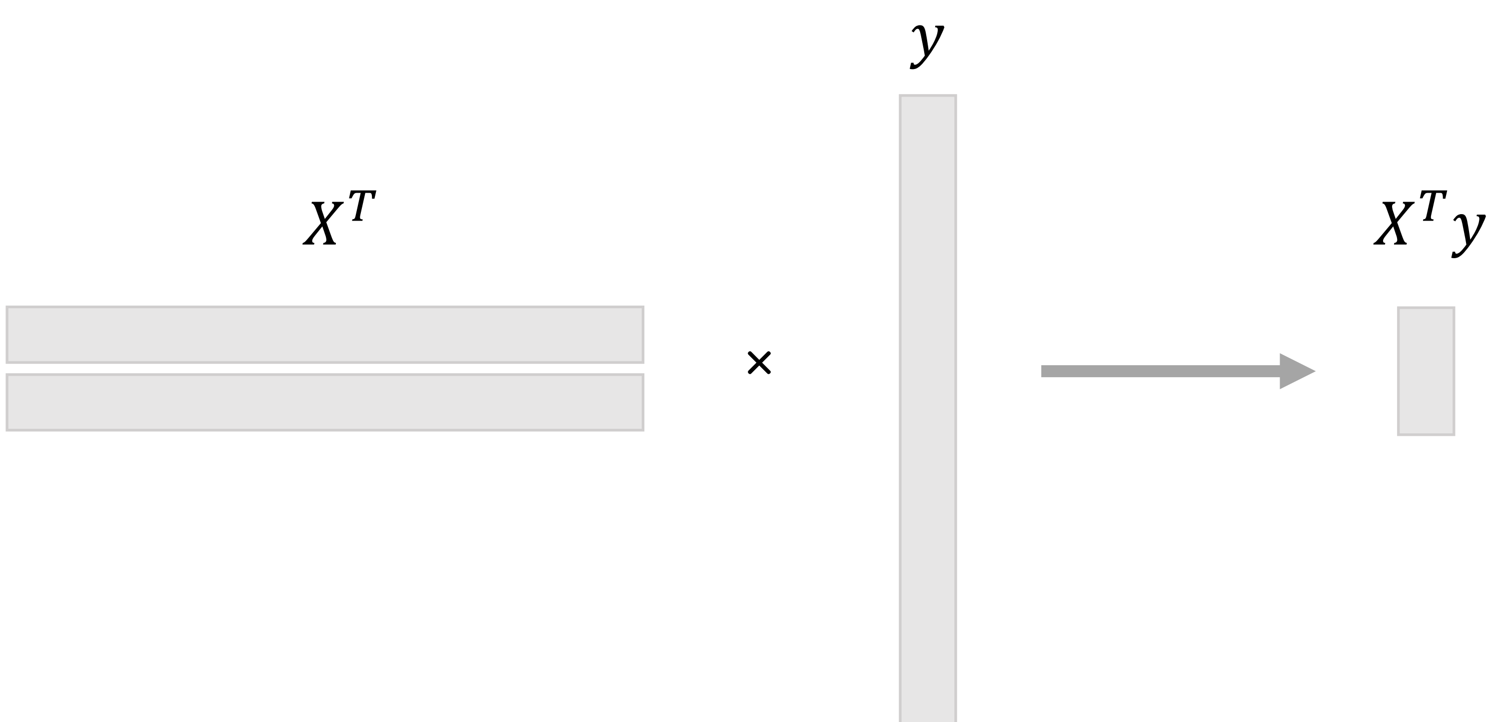
もしyが分かっていたら？

- $X\beta^T$がyに最も近似するβを知りたい
- $\frac{1}{2}(X\beta^T - y)^2$ が最小となるβを計算したい
- βで偏微分して(=0)と置くことで以下を得る

$$\hat{\beta} = (X^T X)^{-1}(X^T y)$$

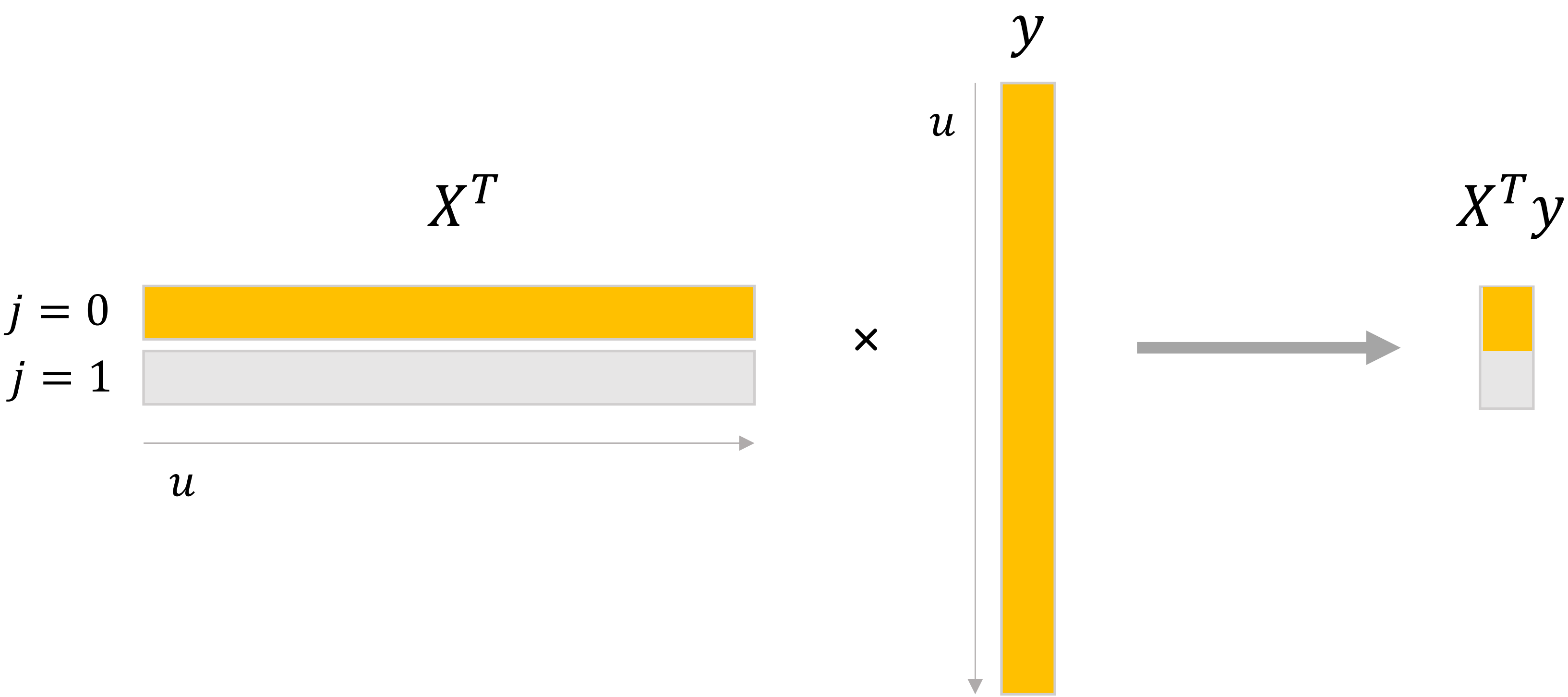
実際yは分からない、けれど

$$\hat{\beta} = (X^T X)^{-1}(X^T y)$$



$X^T y$のj番目の要素を考える

$$x_{j,0}y_0 + x_{j,1}y_1 + \cdots + x_{j,n-1}y_{n-1} = \sum_u x_{ju}y_u$$



式変形

$$\sum_u x_{ju} y_u = \frac{1}{2}\left[\sum_u y_u^2 + \sum_u x_{ju}^2 - \sum_u \left(y_u - x_{ju}\right)^2\right]$$

第1項

$$\sum_u x_{ju} y_u = \frac{1}{2}\left[\sum_u y_u^2 + \sum_u x_{ju}^2 - \sum_u \left(y_u - x_{ju}\right)^2\right]$$

- 全ての値を0で提出した際の最小二乗誤差
- public LBのスコアで近似

第2項

$$\sum_u x_{ju} y_u = \frac{1}{2}\left[\sum_u y_u^2 + \sum_u x_{ju}^2 - \sum_u \left(y_u - x_{ju}\right)^2\right]$$

- Xは計算可能

第3項

$$\sum_u x_{ju} y_u = \frac{1}{2}\left[\sum_u y_u^2 + \sum_u x_{ju}^2 - \sum_u \left(y_u - x_{ju}\right)^2\right]$$

- 最小二乗誤差
- public LBのスコアで近似

整理すると

- $\sum_u x_{ju} y_u = \frac{1}{2}\left[\sum_u y_u^2 + \sum_u x_{ju}^2 - \sum_u \left(y_u - x_{ju}\right)^2\right]$ は近似計算できる
- よって $X^T y$ が分かる
- すなわち $\hat{\beta} = (X^T X)^{-1}(X^T y)$ が分かる

まとめ

- Linear Quiz Blending を解説した
- public LBの情報を用いて、優れたensembleの重みを近似計算する
- 評価指標が最小二乗誤差の場合に利用できる
- 必要な情報は
  1. 複数の(予測値, public LBのスコア)のペア
  2. 全て0で提出した際のpublic LBのスコア