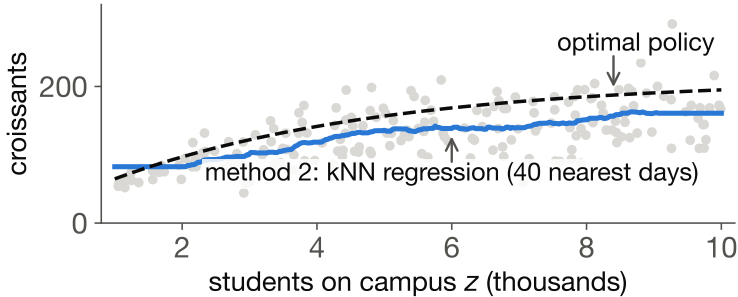
Predict, then optimize

208424 Lecture 14

Lecture 13: the predicted average

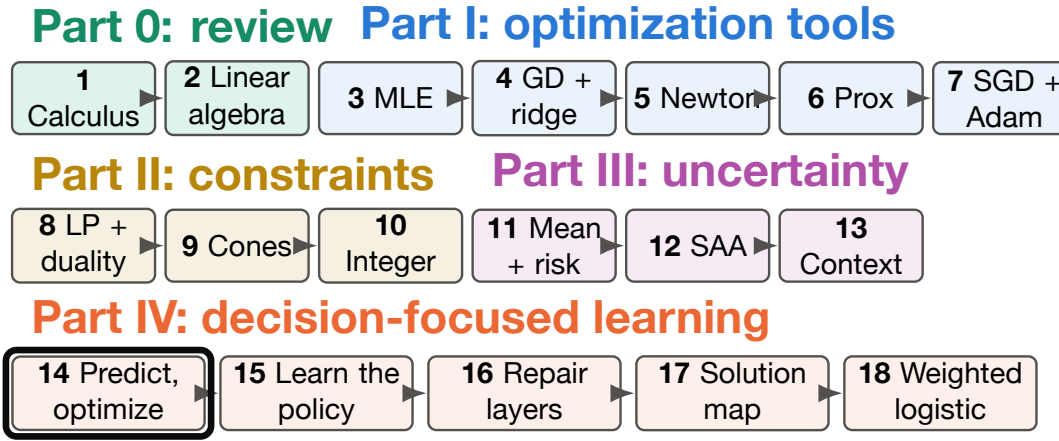


In Lecture 13, method 2 ordered the average.

Method 2 lost 242 baht per day.

When does a good prediction give a good decision?

Course map


Part 0: review
Part I: optimization tools
1 Calculus
2 Linear algebra
3 MLE
4 GD + ridge
5 Newton
6 Prox
7 SGD + Adam
Part II: constraints
Part III: uncertainty
8 LP + duality
9 Cones
10 Integer
11 Mean + risk
12 SAA
13 Context
Part IV: decision-focused learning
14 Predict, optimize
15 Learn the policy
16 Repair layers
17 Solution map
18 Weighted logistic

Lecture 14: Predict, then optimize

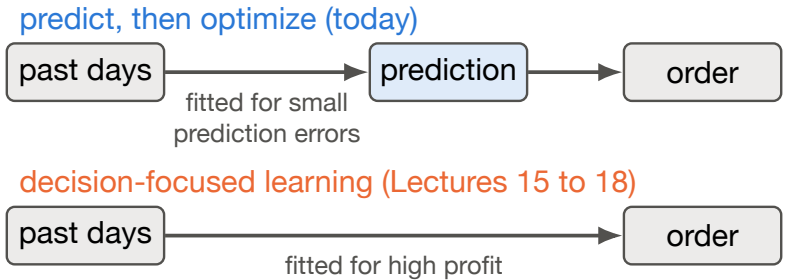


Photo: Love Krittaya, public domain, Wikimedia Commons

A rider predicts how long each road takes.

A café predicts tomorrow's demand.

Part IV: decision-focused learning



Part IV is called decision-focused learning.

A model is fitted for the profit of its decisions.

Today, we study predict, then optimize first.

Medicines in Sierra Leone



Photo: Mattes, public domain, Wikimedia Commons (cropped)

Chung et al. (2022)

With decision-focused learning,
unmet demand was 1.3% in a test on past data.
With the existing tool, unmet demand was 15%.

Inventory at Amazon

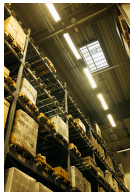
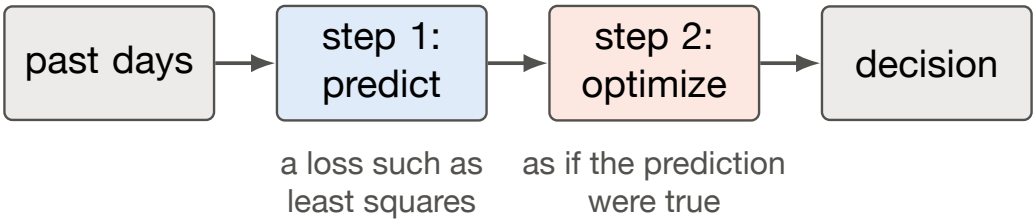


Photo: Shixart1985, CC BY 2.0, Wikimedia Commons

Madeka et al. (2022)

With decision-focused learning,
inventory was 12% lower than with the old policy.
Revenue showed no clear change.

Predict, then optimize

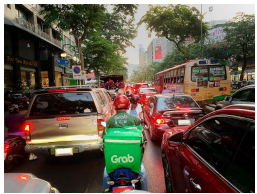


Step 1 fits a prediction to the past days.

Step 2 optimizes as if the prediction were true.

Method 2 of Lecture 13 used these two steps.

A delivery rider



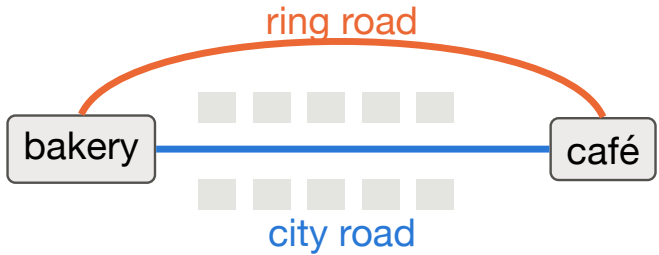


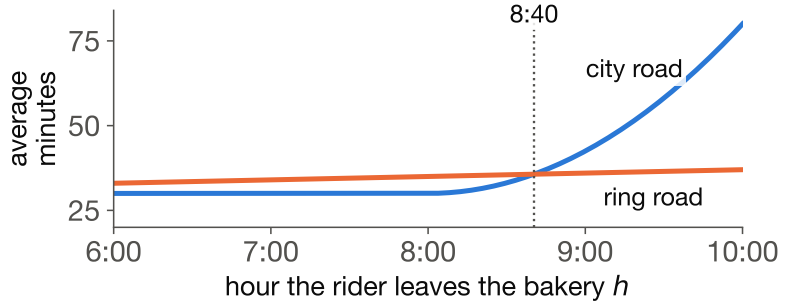
Photo: Chainwit., CC BY-SA 4.0, Wikimedia Commons

A rider brings croissants to the café.

The rider takes the city road or the ring road.

The rider leaves the bakery at hour h.

The average travel time on each road

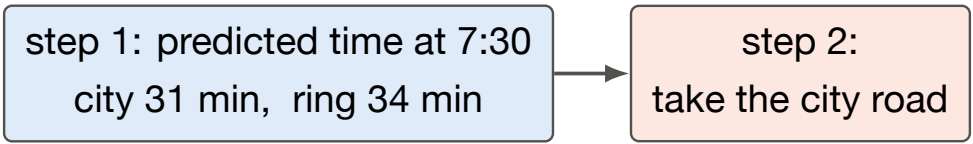


Before 8:00, the city road takes 30 minutes.

After 8:00, the city road gets slower quickly.

The city road is faster until 8:40.

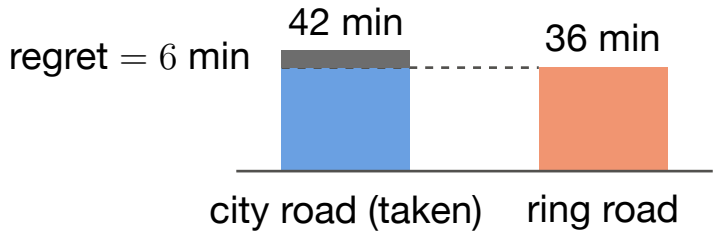
Predict, then optimize for the rider



Step 1 predicts the time on each road at hour h.

Step 2 takes the road with the smaller prediction.

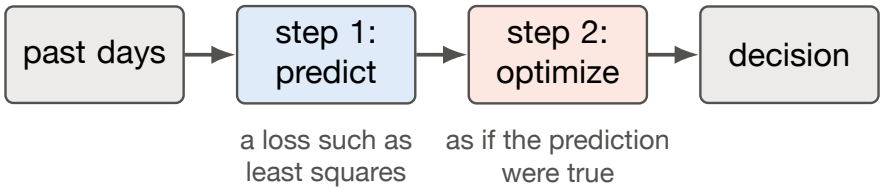
Regret



The regret of a trip is the time on the chosen road minus the time on the faster road for that trip.

On this trip, the regret is $42 - 36 = 6$ minutes.

Summary: predict, then optimize



Step 2 optimizes as if the prediction were true.

Regret is the time lost on one trip.

Checkpoint 1

predicted:	actual:
city 30 min, ring 35 min	city 40 min, ring 33 min

The rider takes the road with the smaller prediction.

What is the regret of this trip?

Checkpoint 1

predicted:	actual:
city 30 min, ring 35 min	city 40 min, ring 33 min

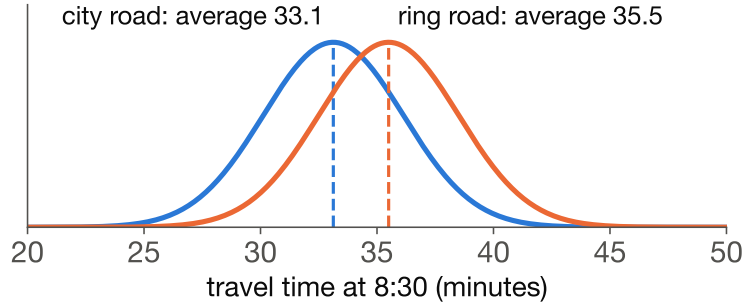
The rider takes the road with the smaller prediction.

What is the regret of this trip?

The rider takes the city road, since $30 < 35$.

The regret is $40 - 33 = 7$ minutes.

What should the rider predict?



At 8:30, the travel time varies from day to day.

The rider wants the smaller average time.

So step 1 should predict the average time.

Road A and road B

road A	30	30	30	30	30	30	30	30	30	80
road B	33	33	33	33	33	33	33	33	33	33

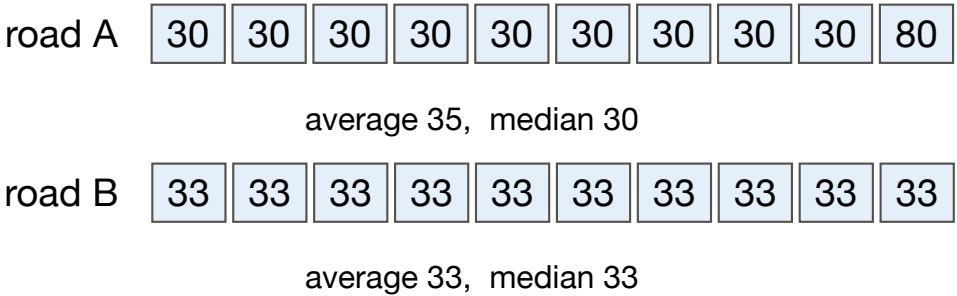
minutes on 10 days

Road A takes 30 minutes on 9 days out of 10.

On the other day, road A takes 80 minutes.

Road B always takes 33 minutes.

The average or the median?

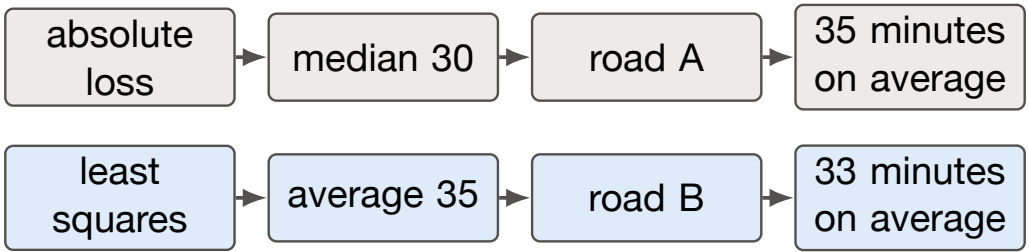


Road A has an average of 35 and a median of 30.

With median predictions, the rider takes road A.

With average predictions, the rider takes road B.

Which loss gives the faster road?



The absolute loss fits the median (Lecture 3).

Least squares fits the average (Lecture 13).

For the rider, least squares is the right loss.

The café needs a quantile

one croissant too few	one croissant too many
-60 baht	-20 baht

$$60/(60+20) = 0.75$$

One croissant too few loses a profit of 60 baht.

One croissant too many loses its cost of 20 baht.

So the order is the 0.75 quantile of demand.

Summary: what to predict

rider:
predict the
average time

café:
predict the distribution
of demand

Step 1 must predict what the decision needs.

For the café, the average is not enough.

Checkpoint 2

road C	20	20	20	20	20	20	20	20	60	60
road D	26	26	26	26	26	26	26	26	26	26

Step 1 predicts the median time of each road.

Does the rider choose the faster road on average?

Checkpoint 2

road C	20	20	20	20	20	20	20	20	60	60
road D	26	26	26	26	26	26	26	26	26	26

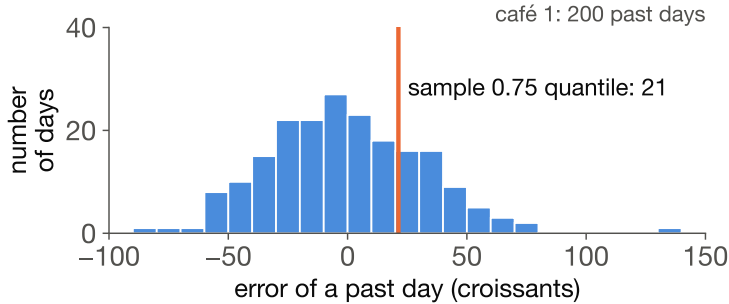
Step 1 predicts the median time of each road.

Does the rider choose the faster road on average?

No. With medians, the rider takes road C.

But road C takes 28 minutes on average.

The errors of the predicted average

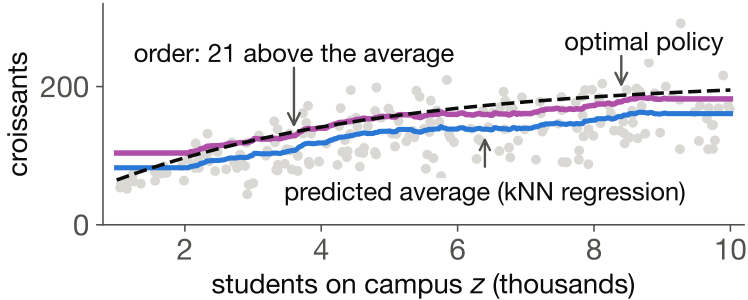


Now step 1 predicts the distribution of demand.

The error of day i is D_i minus the predicted average.

The sample 0.75 quantile of the errors is 21.

Adding the error quantile

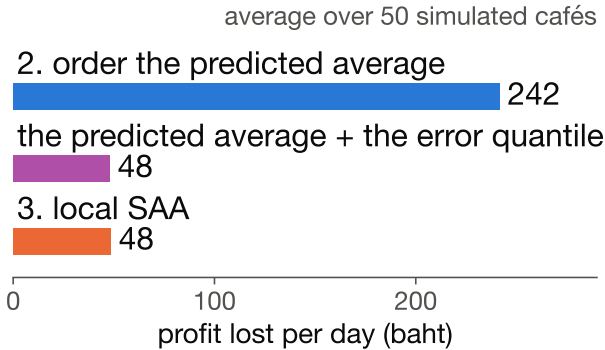


Step 2 orders 21 croissants above the average.

About 75% of past days had enough croissants.

This order is close to the optimal policy.

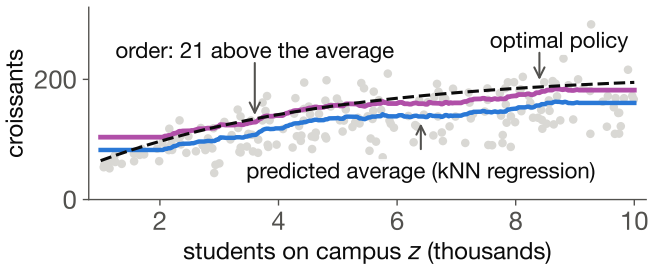
Profit lost with the error quantile



The profit lost falls from 242 to 48 baht per day.

This method is as good as local SAA.

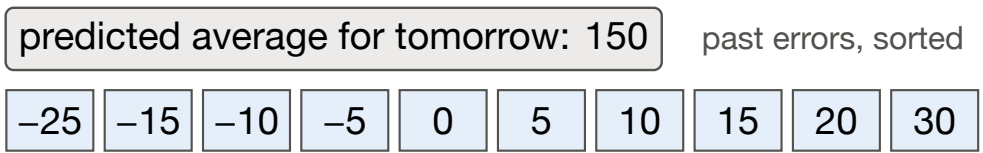
Summary: the error quantile



Step 1 also finds the errors of past days.

Step 2 adds the 0.75 quantile of the errors.

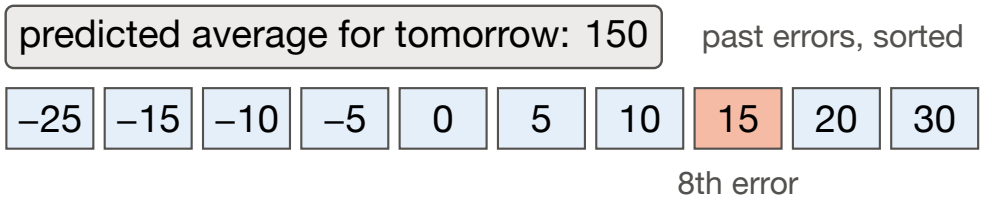
Checkpoint 3



The cost is 20 baht, and the price is 80 baht.

How many croissants should the café order?

Checkpoint 3



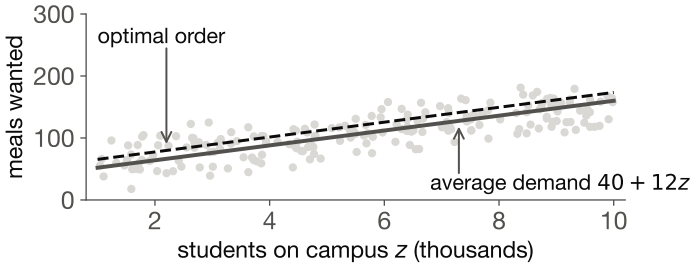
The cost is 20 baht, and the price is 80 baht.

How many croissants should the café order?

Since $0.75 \times 10 = 7.5$, we take the 8th error.

So the café orders $150 + 15 = 165$ croissants.

A canteen on campus



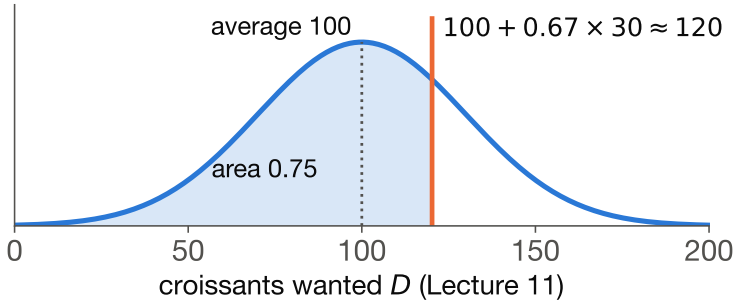
cost $c = 20$ baht
price $q = 80$ baht

A canteen sells meals to the students on campus.

The average demand is $40 + 12z$ meals.

Every day, the demand is normal with an SD of 20.

The 0.75 quantile of normal demand



For a normal distribution, the 0.75 quantile is 0.67 SD above the average.

Lecture 11 found the same order of 120.

Two methods for ordering meals

predict, then optimize

order $=$ line $+ 0.67$ SD

quantile regression
(Lecture 13)

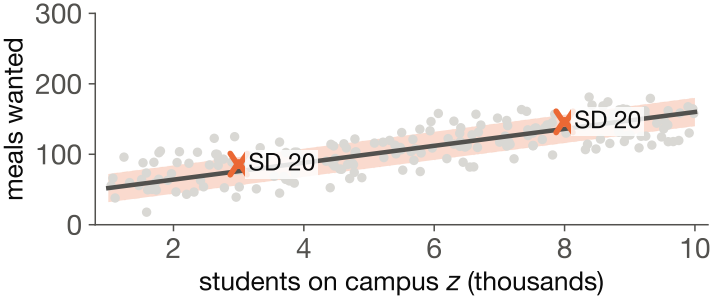
order $=$ 0.75 quantile line

Step 1 fits a line and the SD of its errors.

Step 2 orders 0.67 SD above the line.

Quantile regression fits the 0.75 quantile line.

A well-specified model

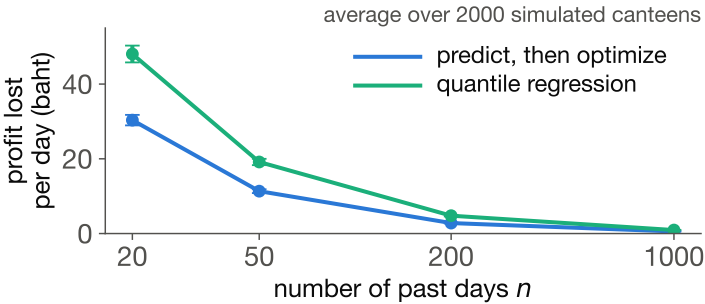


Step 1's model uses a line and one SD for all days.

The canteen's demand has exactly this form.

So the model is called well specified.

Profit lost as the number of days grows

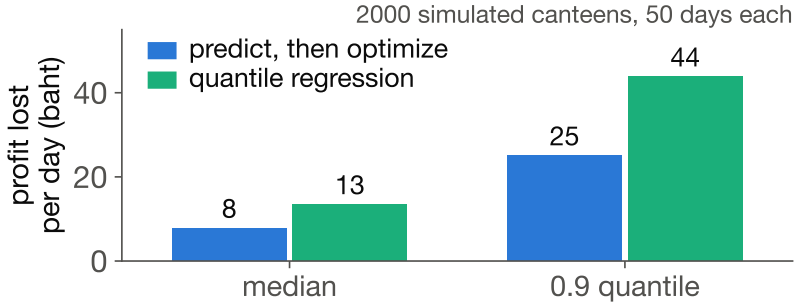


Predict, then optimize loses less for every n.

With 50 days, the two losses are 11 and 19 baht.

The optimal profit is 5852 baht per day.

Profit lost in 2000 canteens

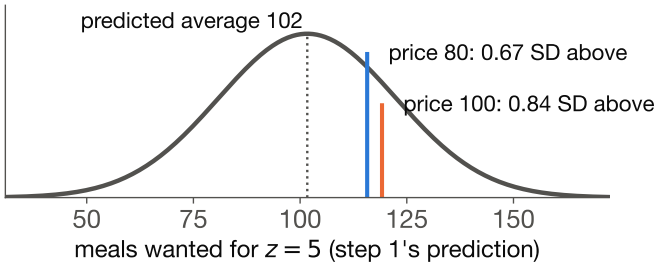


We simulated 2000 canteens with 50 days each.

Predict, then optimize has a lower median loss.

The 0.9 quantile of the loss is also lower.

A new price



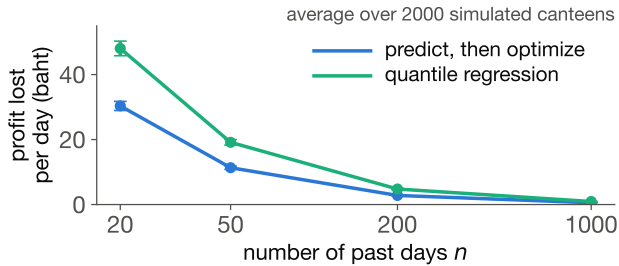
price 100 baht:
$(100 - 20)/100 = 0.80$

At a price of 100 baht, we order the 0.80 quantile.

Step 2 orders 0.84 SD above the average.

Quantile regression must be fitted again.

Summary: a well-specified model



With a well-specified model,

predict, then optimize loses less.

The same prediction works for a new price.

Checkpoint 4

predicted demand: normal, average 150, SD 20

cost 20 baht, price 80 baht

How many meals should the canteen order?

Checkpoint 4

predicted demand: normal, average 150, SD 20

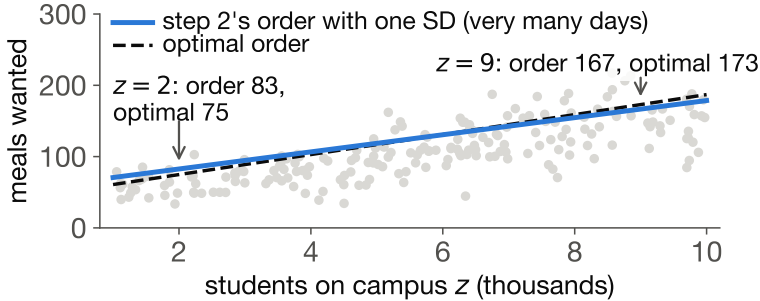
cost 20 baht, price 80 baht

How many meals should the canteen order?

The order is 0.67 SD above the average.

So the order is $150 + 0.67 \times 20 \approx 163$.

The canteen with a growing SD

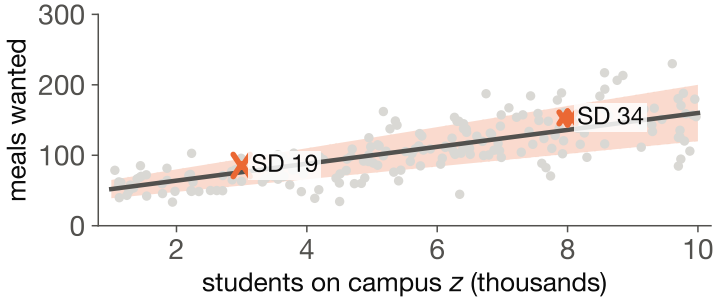


Now the SD grows with the average demand.

For small z, step 2 orders too many meals.

For large z, step 2 orders too few meals.

A misspecified model

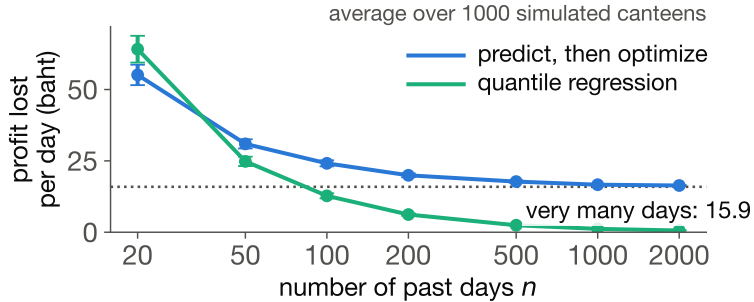


A line with one SD cannot describe this demand.

Such a model is called misspecified.

The optimal order is still a line in z.

Profit lost with a growing SD

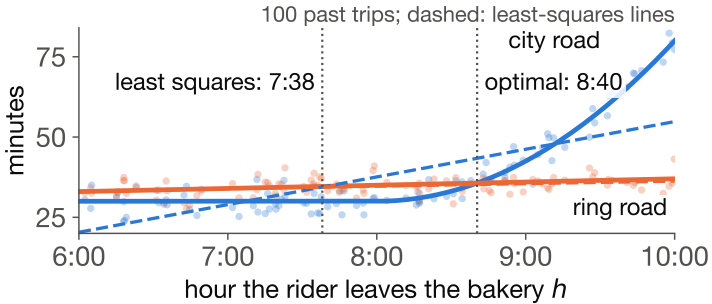


With 20 days, predict, then optimize loses less.

From 50 days on, quantile regression loses less.

At $n = 2000$, predict, then optimize loses 16.4 baht.

Linear predictions for the two roads

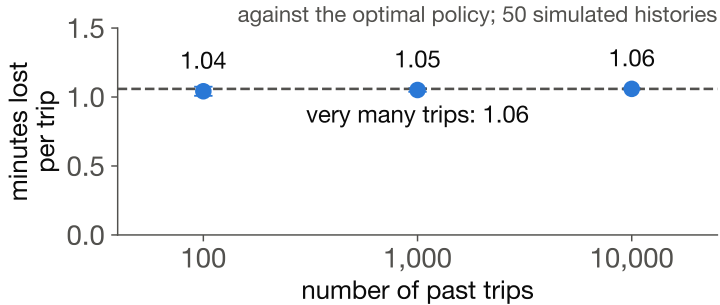


The city road's average time is not linear in h.

So a line is a misspecified model for this road.

With least-squares lines, the rider switches at 7:38.

Minutes lost with more trips

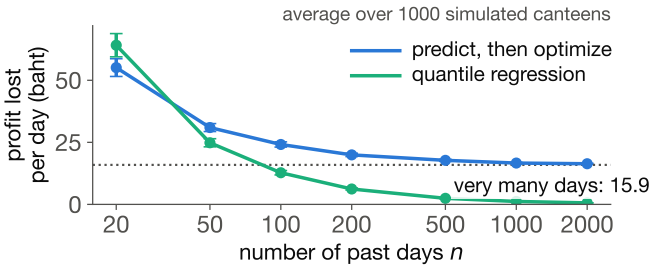


From 7:38 to 8:40, the rider takes the slower road.

With 100 trips, the rider loses 1.04 minutes per trip.

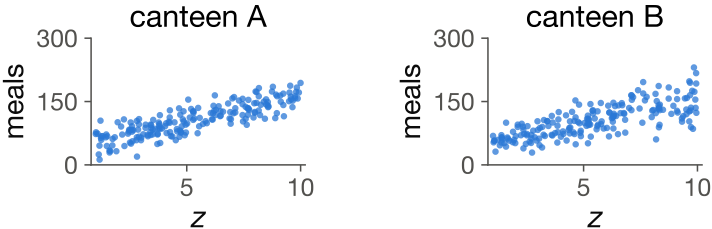
With 10,000 trips, the rider still loses 1.06 minutes.

Summary: a misspecified model



With a misspecified model, predict, then optimize still loses with very many days.

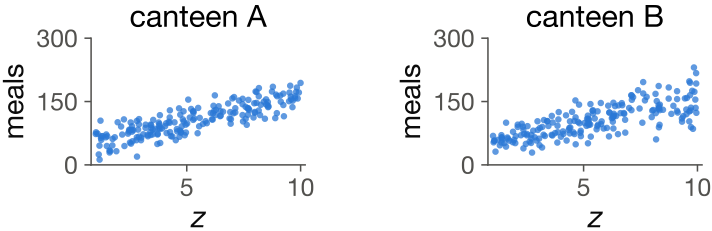
Checkpoint 5



Step 1's model uses a line and one SD for all days.

For which canteen is the model well specified?

Checkpoint 5



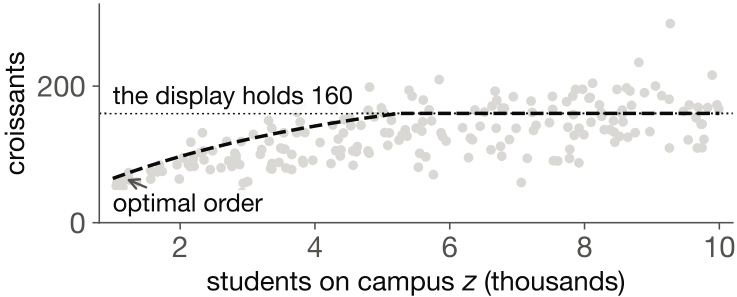
Step 1's model uses a line and one SD for all days.

For which canteen is the model well specified?

The model is well specified for canteen A.

In canteen B, the SD grows with z.

A display that holds 160 croissants

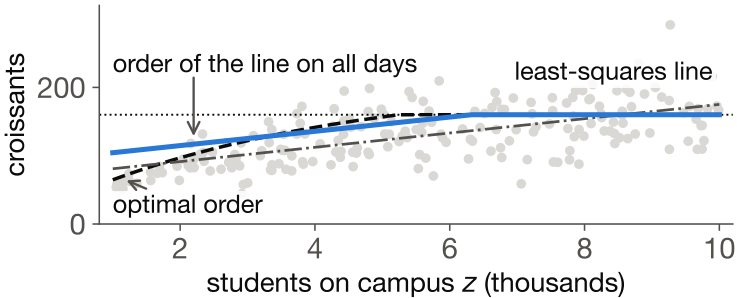


The café of Lecture 13 now has a display.

The display holds at most 160 croissants.

For large z, the optimal order is 160.

A line for the café

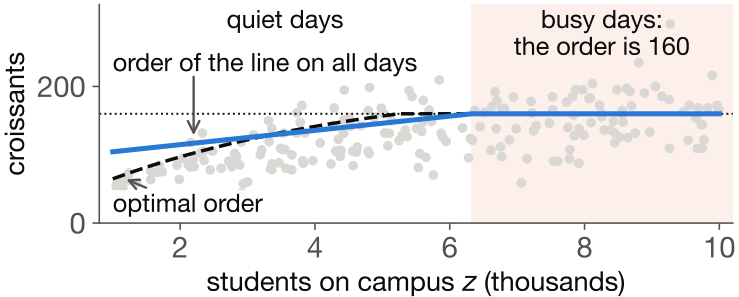


Step 1 fits a least-squares line to all 200 days.

Step 2 adds the 0.75 quantile of the errors.

The average demand is not linear in z.

Busy days change the line

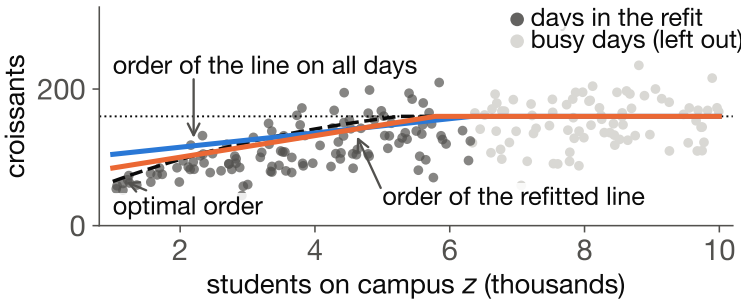


Least squares fits one line to all 200 days.

So busy days also change the line on quiet days.

kNN regression uses only the nearest days.

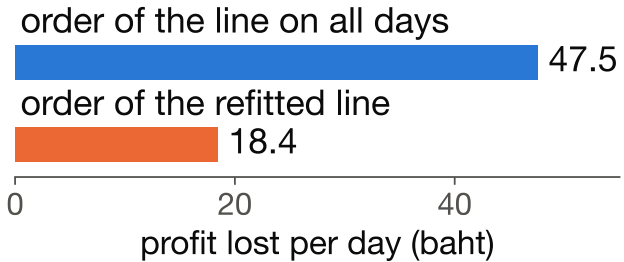
Refitting the line (Chung et al., 2022)



On busy days, the order is always 160.

So we fit the line again without the busy days.

Profit lost with the refitted line

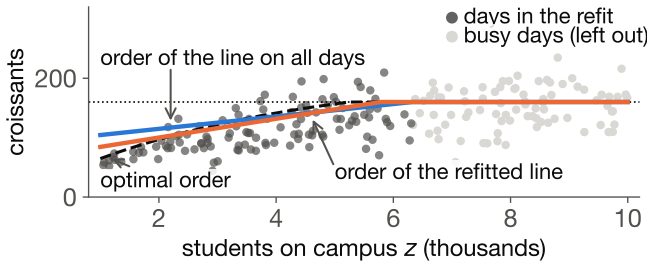


We simulated 50 cafés with a display of 160.

The loss falls from 47.5 to 18.4 baht per day.

The refitted line loses less for 49 of 50 cafés.

Summary: a display that holds 160



On busy days, the order is always 160.

Refitting without the busy days lowered the loss.

Checkpoint 6

four days

$z = 3$ | $z = 5$ | $z = 7$ | $z = 9$

the display holds 160

With the line on all days, step 2 orders $40 + 20z$.

Which of these days are in the refit?

Checkpoint 6

four days

$z = 3$ | $z = 5$ | $z = 7$ | $z = 9$

the display holds 160

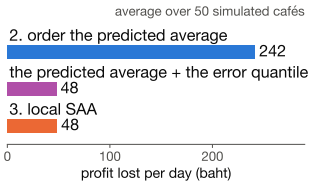
With the line on all days, step 2 orders $40 + 20z$.

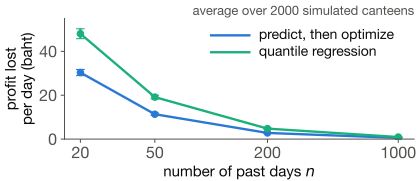
Which of these days are in the refit?

For $z = 7$ and $z = 9$, the order is 160.

So the refit uses the days with $z = 3$ and $z = 5$.

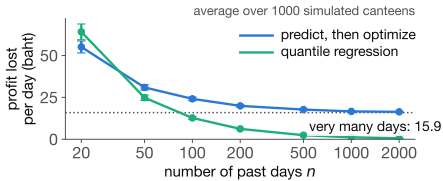
Lecture summary

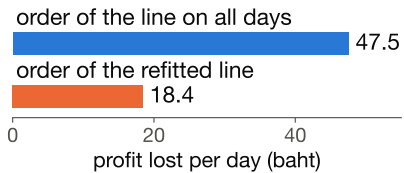




A good prediction gives a good decision when step 1 predicts what the decision needs and when the model is well specified.

Lecture summary (continued)





With a misspecified model, predict, then optimize still loses with very many days.

Refitting without the busy days lowered the loss.

Next lecture

1. ignore z:
SAA (Lecture 12)

2. predict,
then optimize

3. local SAA

4. learn the
policy directly

Quantile regression is method 4 of Lecture 13.

Method 4 learns the policy from past days.

Next, we learn policies by simulating past days.

Exercises

1 road E | 25 | 25 | 25 | 25 | 60 | road F | 30 | 30 | 30 | 30 | 30 |

2 predicted average: 140 past errors | –20 | –10 | 0 | 5 | 10 | 20 |

3 normal demand: average 200, SD 40; price 100 baht
the 0.80 quantile is 0.84 SD above the average

1. Use medians, then averages, to choose a road.
2. Find the order for $c = 20$ and $q = 80$.
3. Find the order for a price of 100 baht.