

Interactive Online Learning

Sanjiban Choudhury



Cornell Bowers CIS
Computer Science

Announcements (all on Ed!)



1. Assignment 0 (survey released)

2. Lecture 1 slides + notes up on website

3. Office hours available:

Sanjiban (Tue/Thurs 11-12pm, Gates 413B)

Dhruv (Mon/Wed 11-12 pm, Rhodes 400)

Learning

Today!

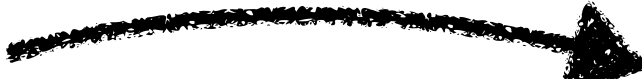
Robot
Decision
Making

Reinforcement Learning



Interactive
Online
Learning

Imitation
Learning



Meta
Learning



Model Predictive
Control



Anytime
Planning



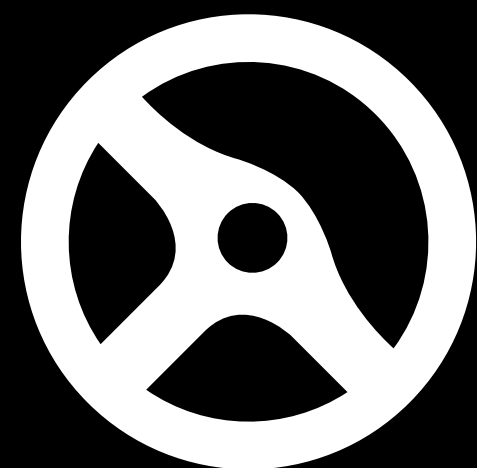
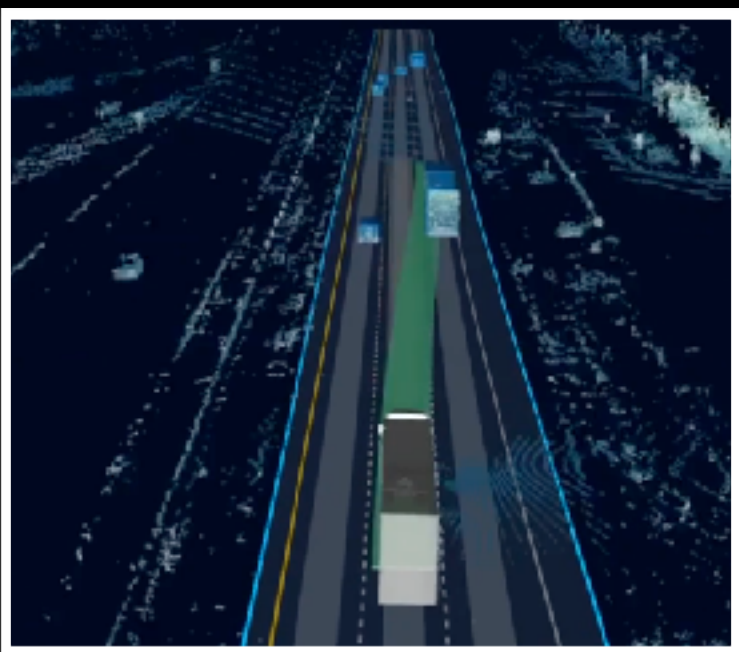
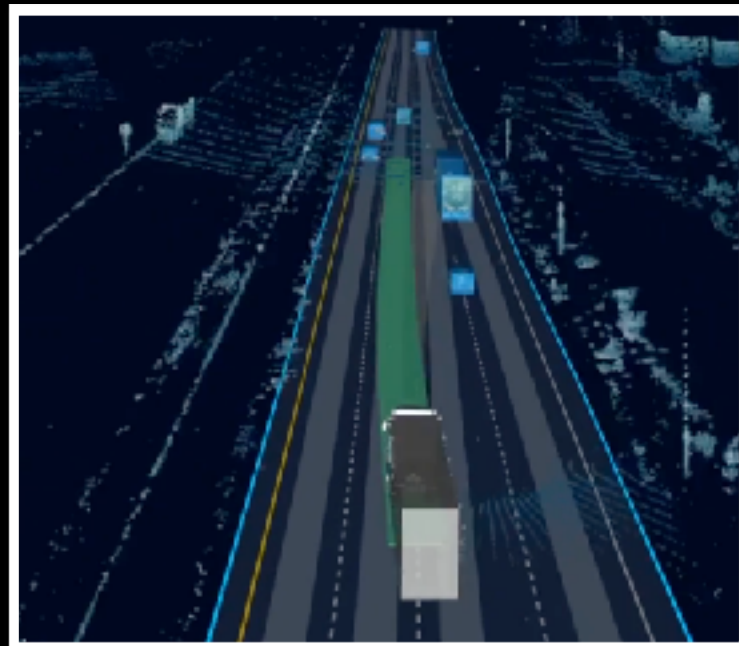
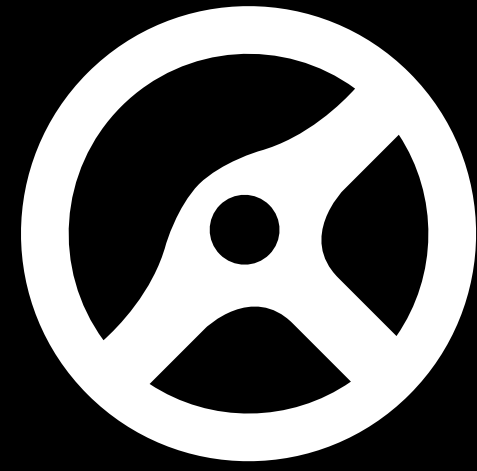
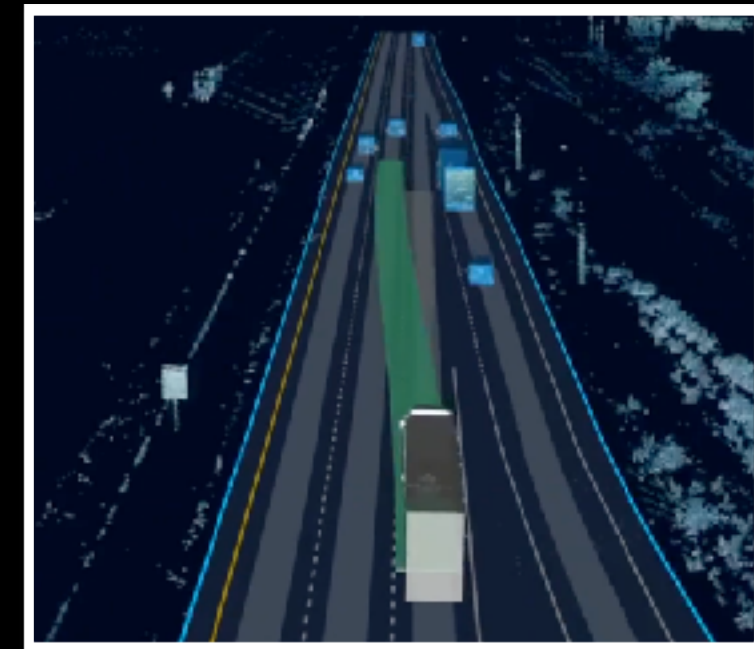
Can't we collect a
LOT of data and
train robots
offline?



SUPERVISED LEARNING

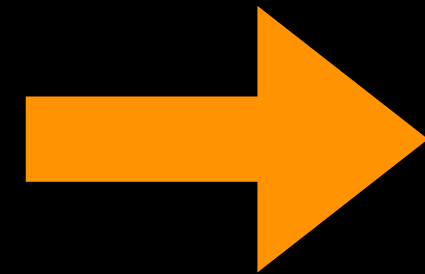


#1 Get Data



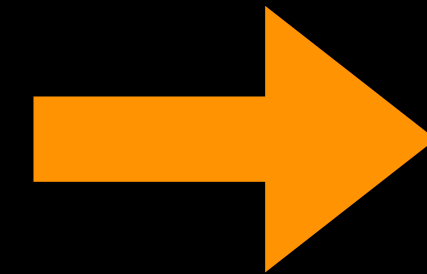
Input (x)

Output (y)



#2
Train
Policy

$$\pi : x \rightarrow y$$



#3 Deploy!



An aerial, high-angle photograph of a dense urban street intersection. The street is filled with a variety of vehicles, including cars, buses, and vans, moving in different directions. The image is dimmed and has a dark, muted color palette. Overlaid on the center of the image is the text "Train ≠ Test" in a large, bold, orange-red font. The text is the primary focus, with the background serving as a visual metaphor for the concept of training and testing in machine learning.

Train \neq Test

Train Test Mismatch

$$\mathbb{E}_{x,y \sim P_{train}(x,y)} \ell(y, \pi(x))$$

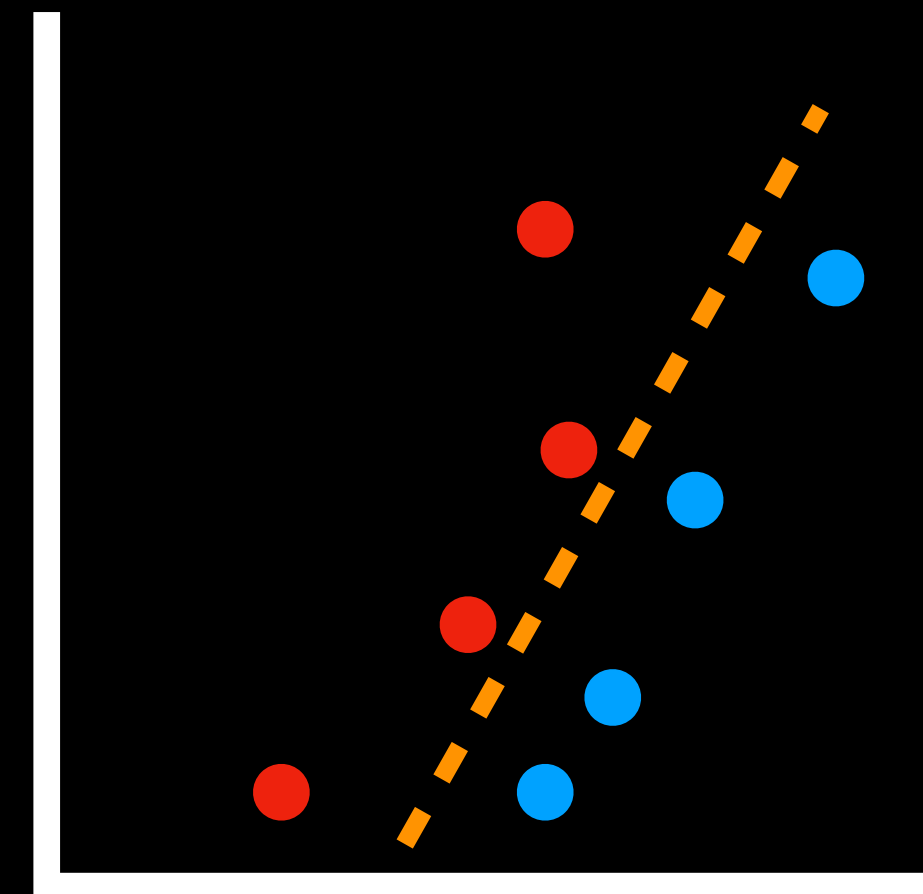
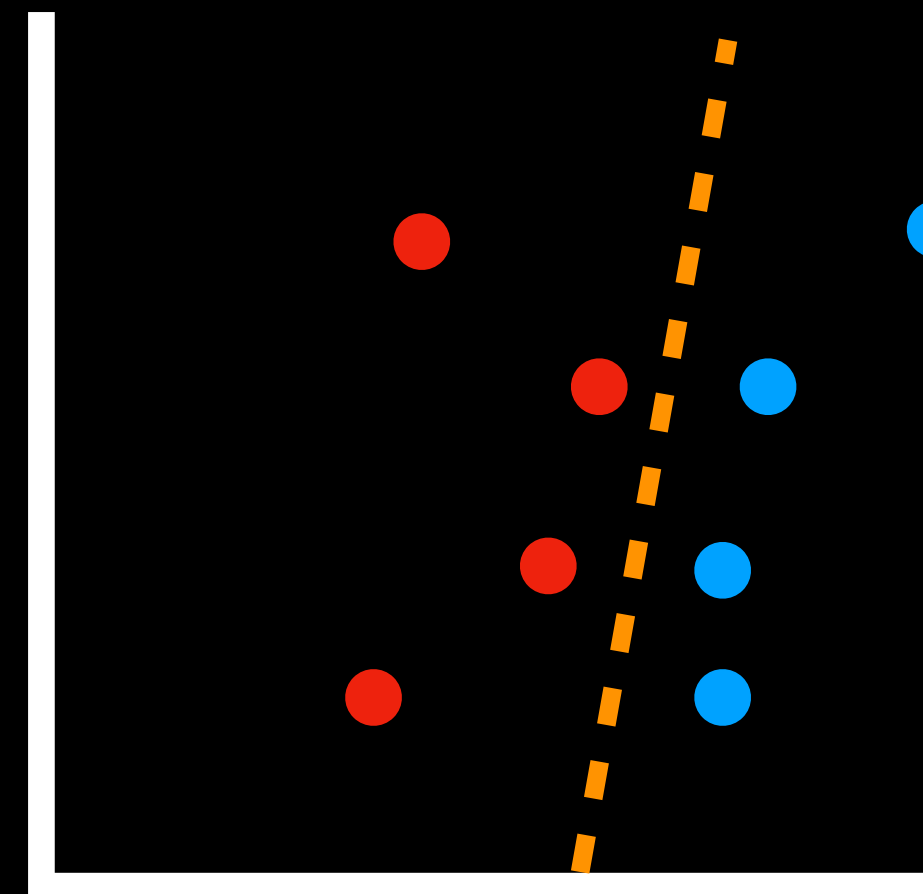
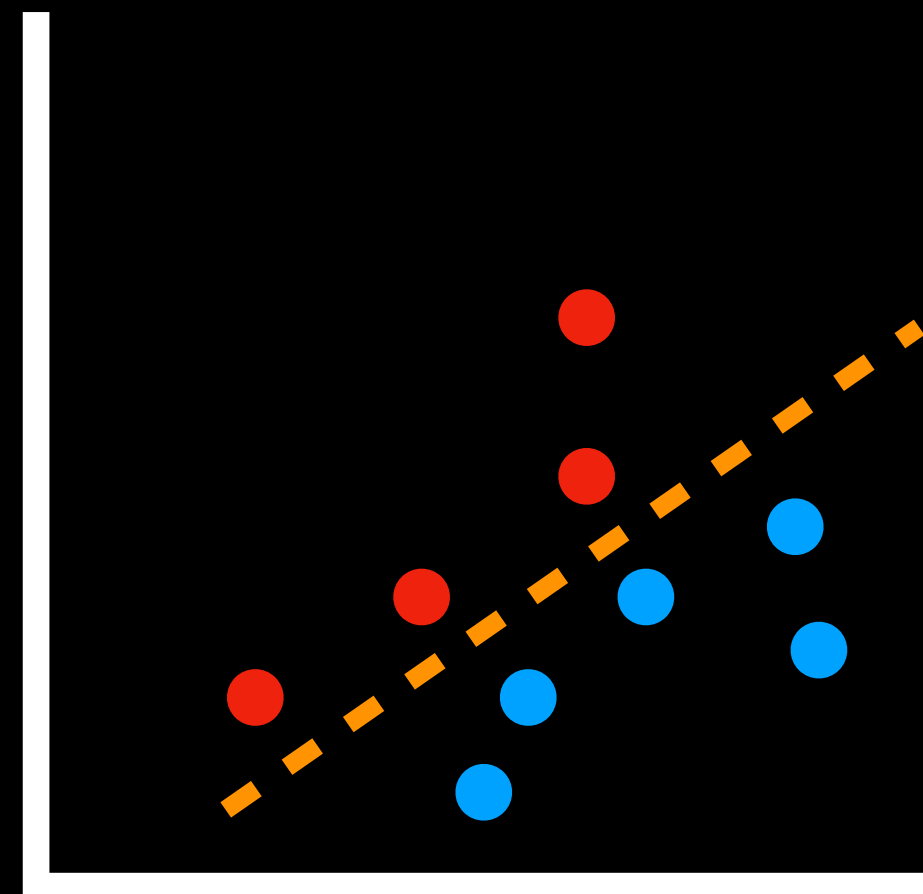
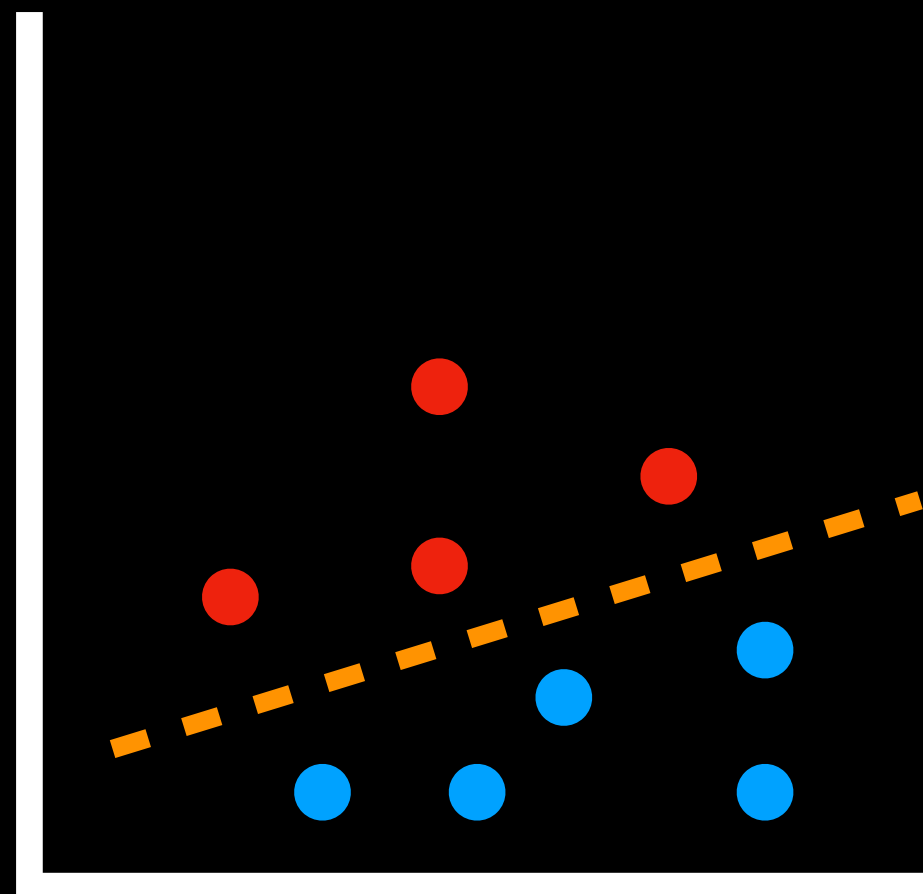
\neq

$$\mathbb{E}_{x,y \sim P_{test}(x,y)} \ell(y, \pi(x))$$

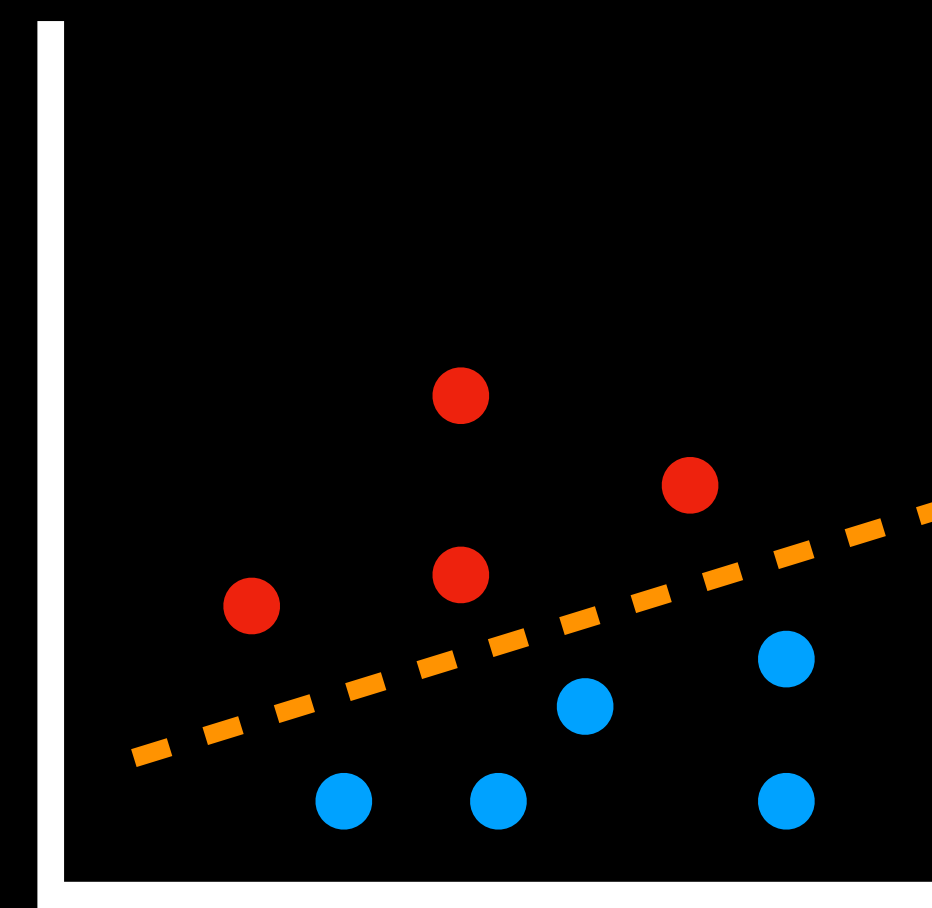
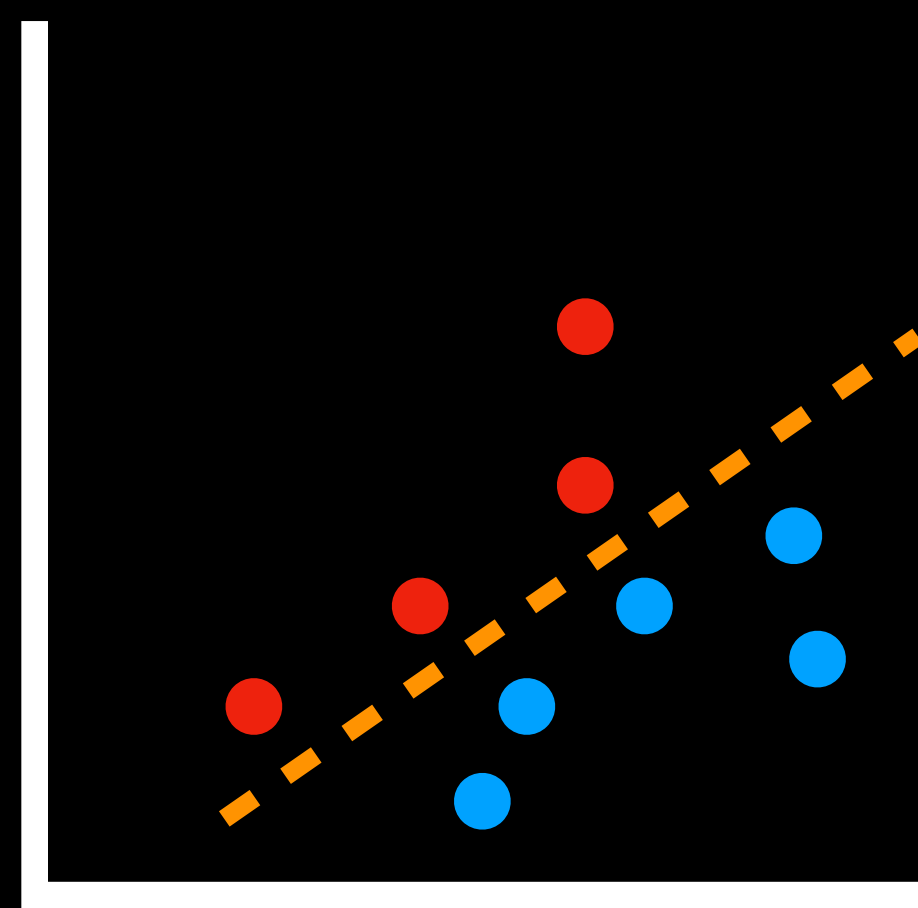
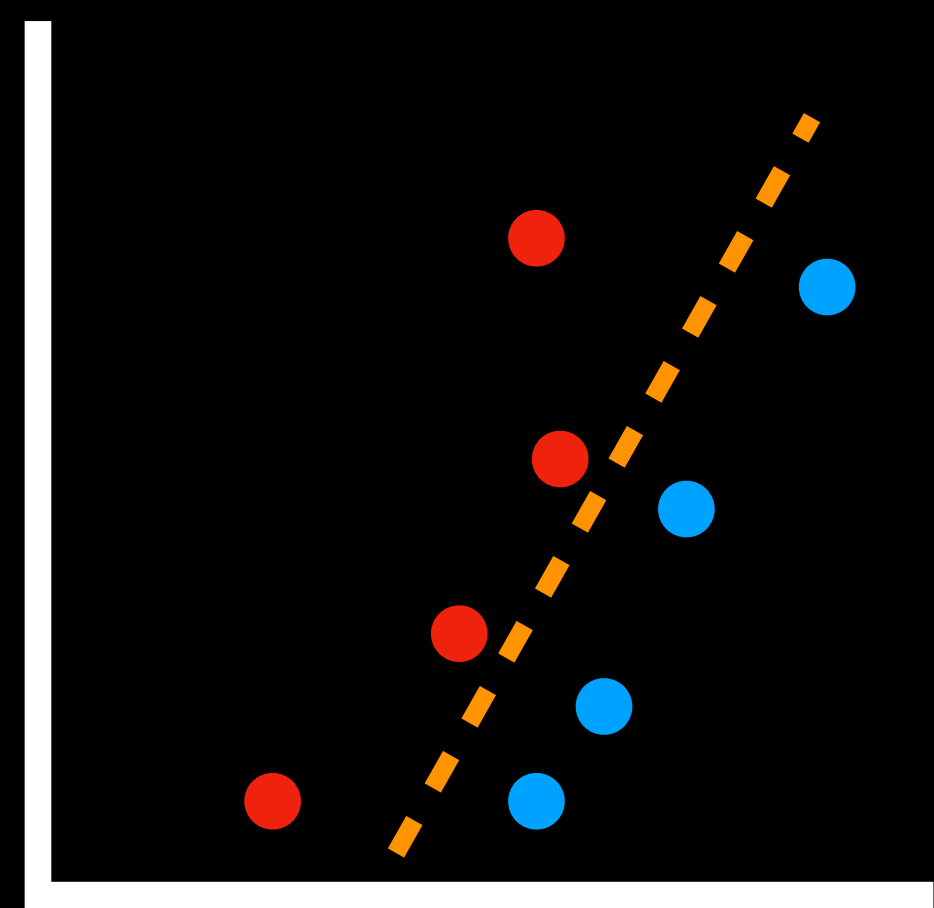
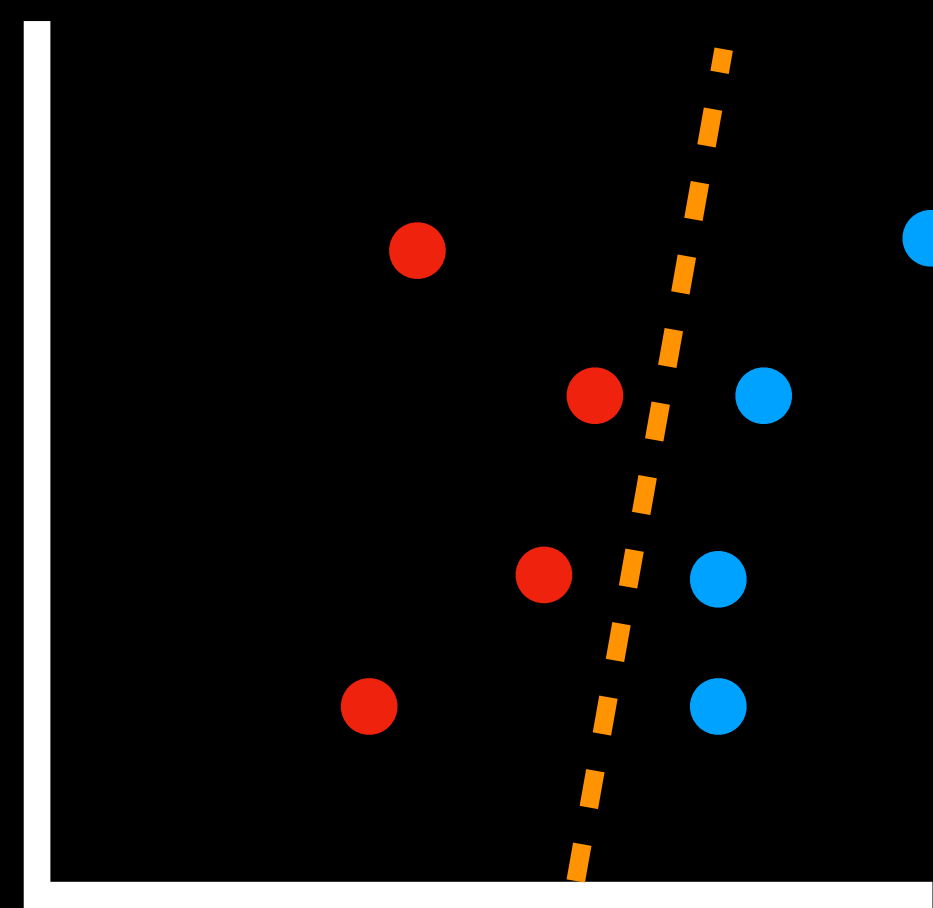
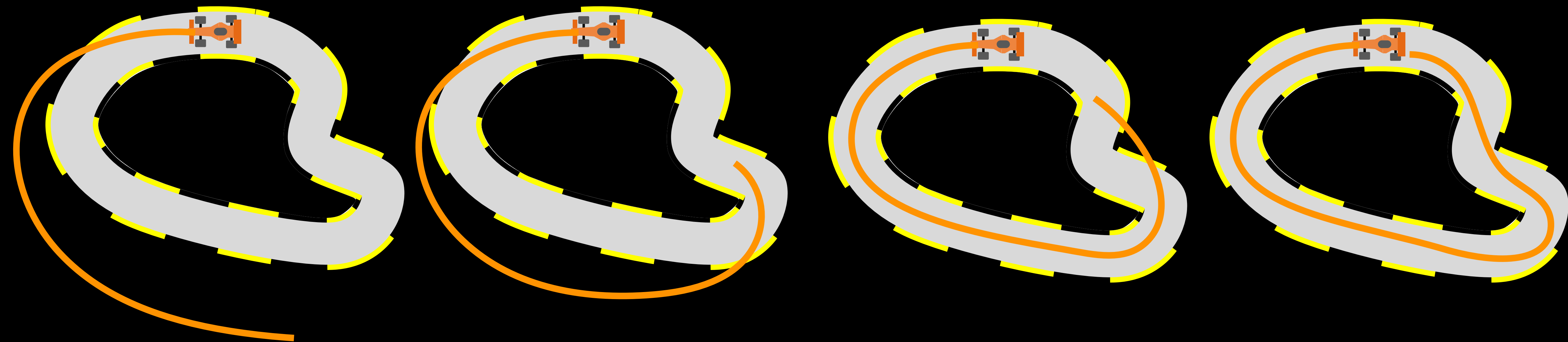
Activity!



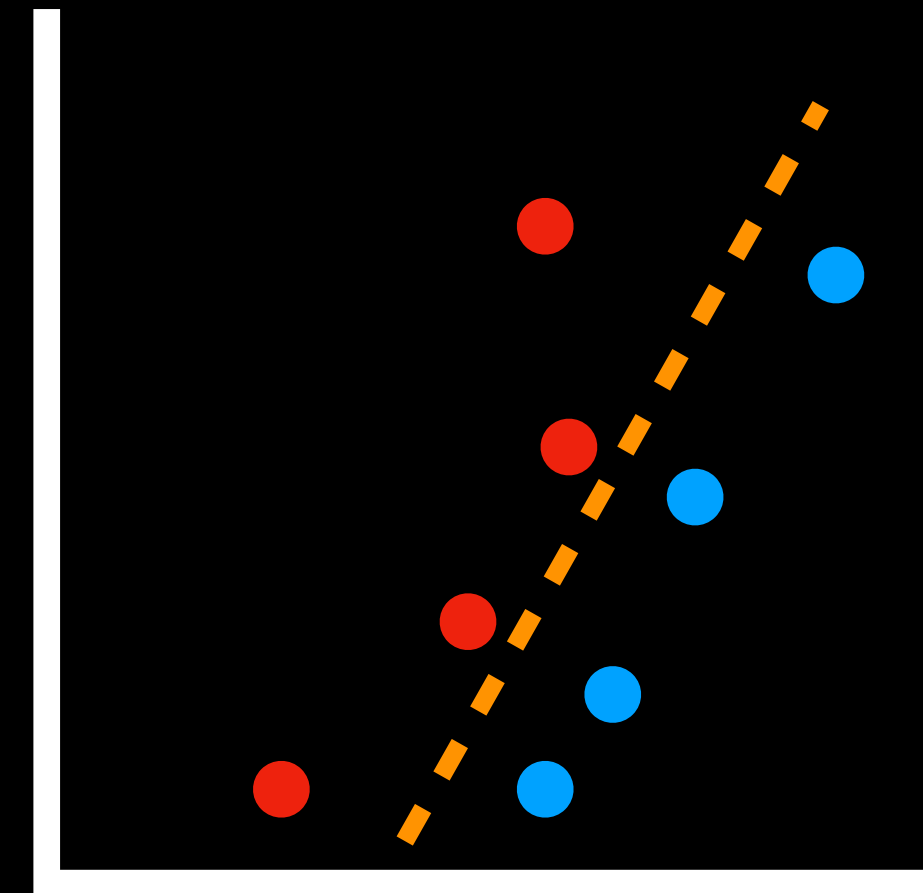
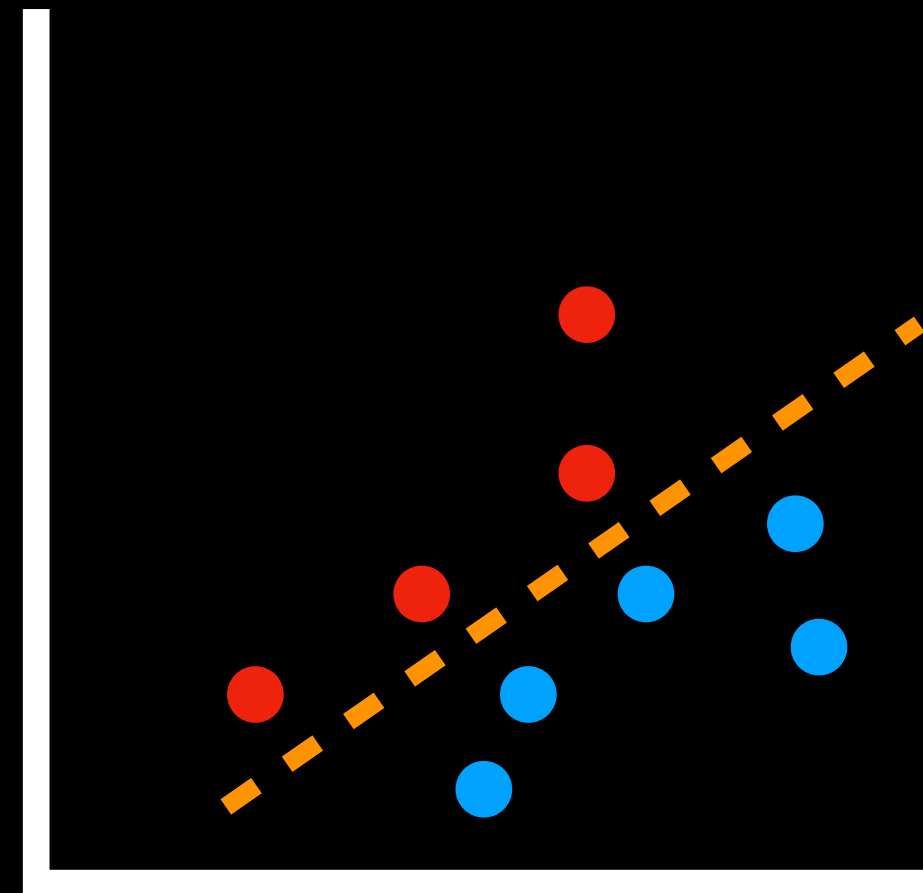
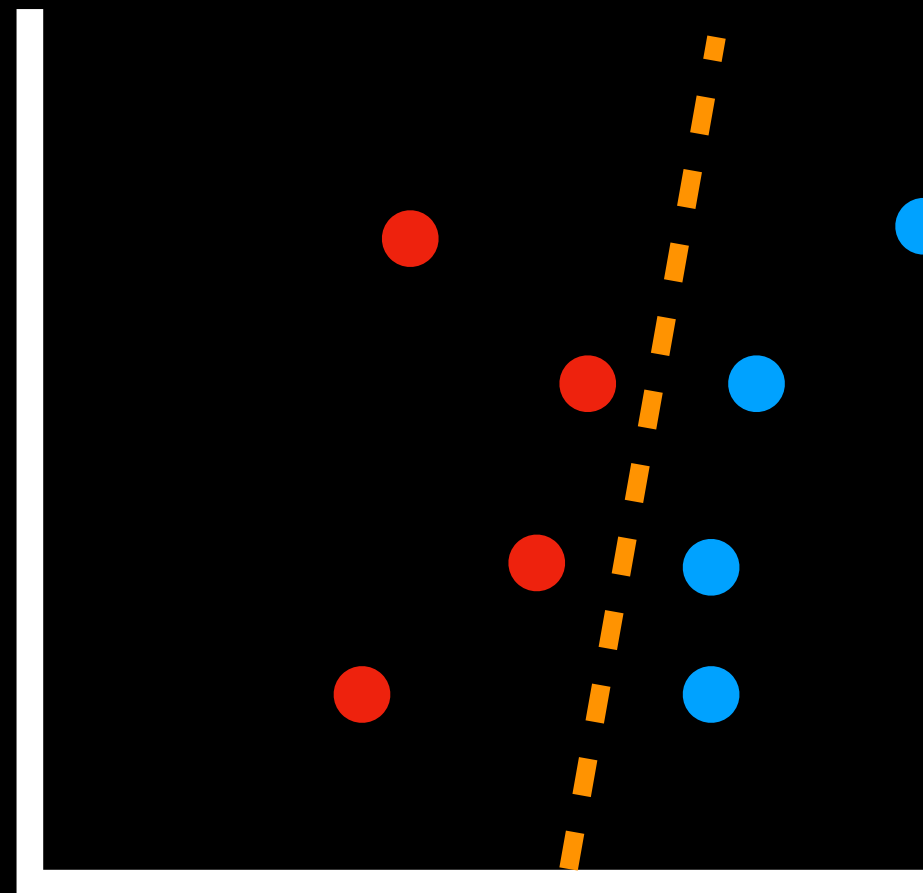
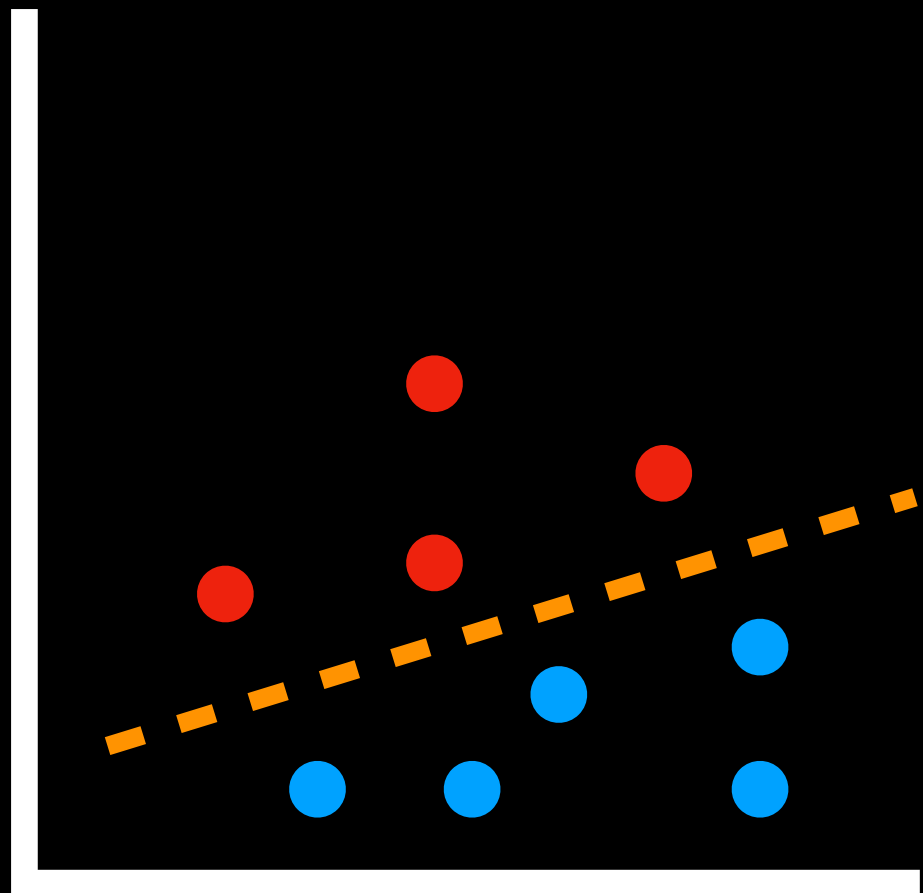
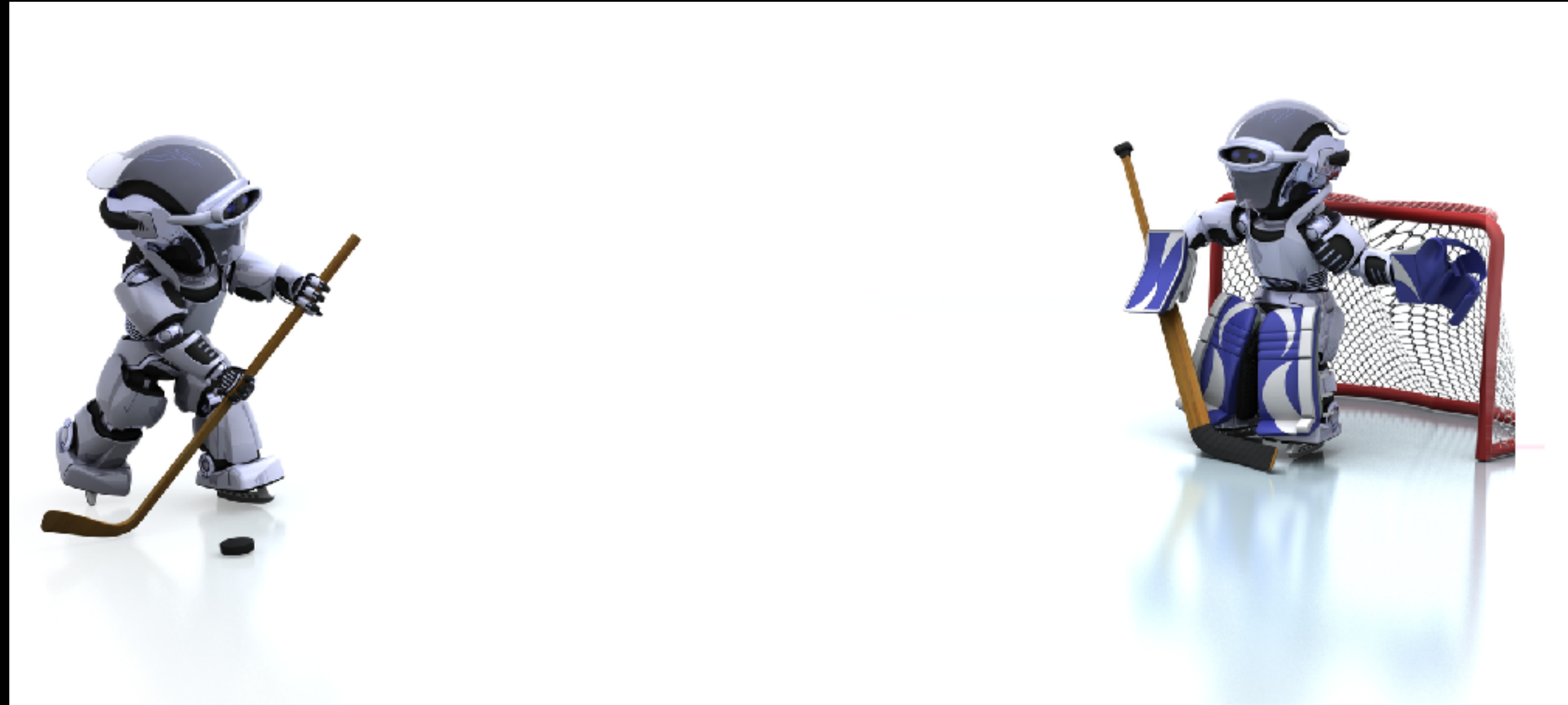
Case 1: $P(x)$ changes over time



Case 2: $P(x)$ influenced by robot π

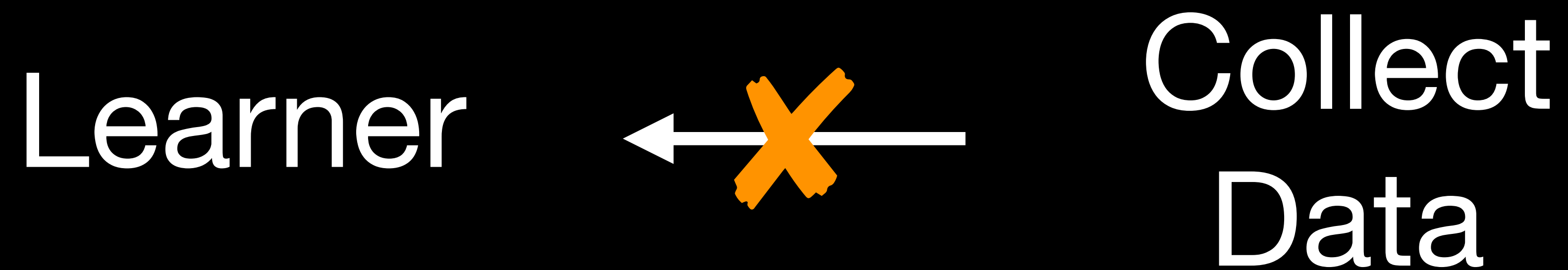


Case 3: $P(x,y)$ changes **adversarially** (game)



Challenge:

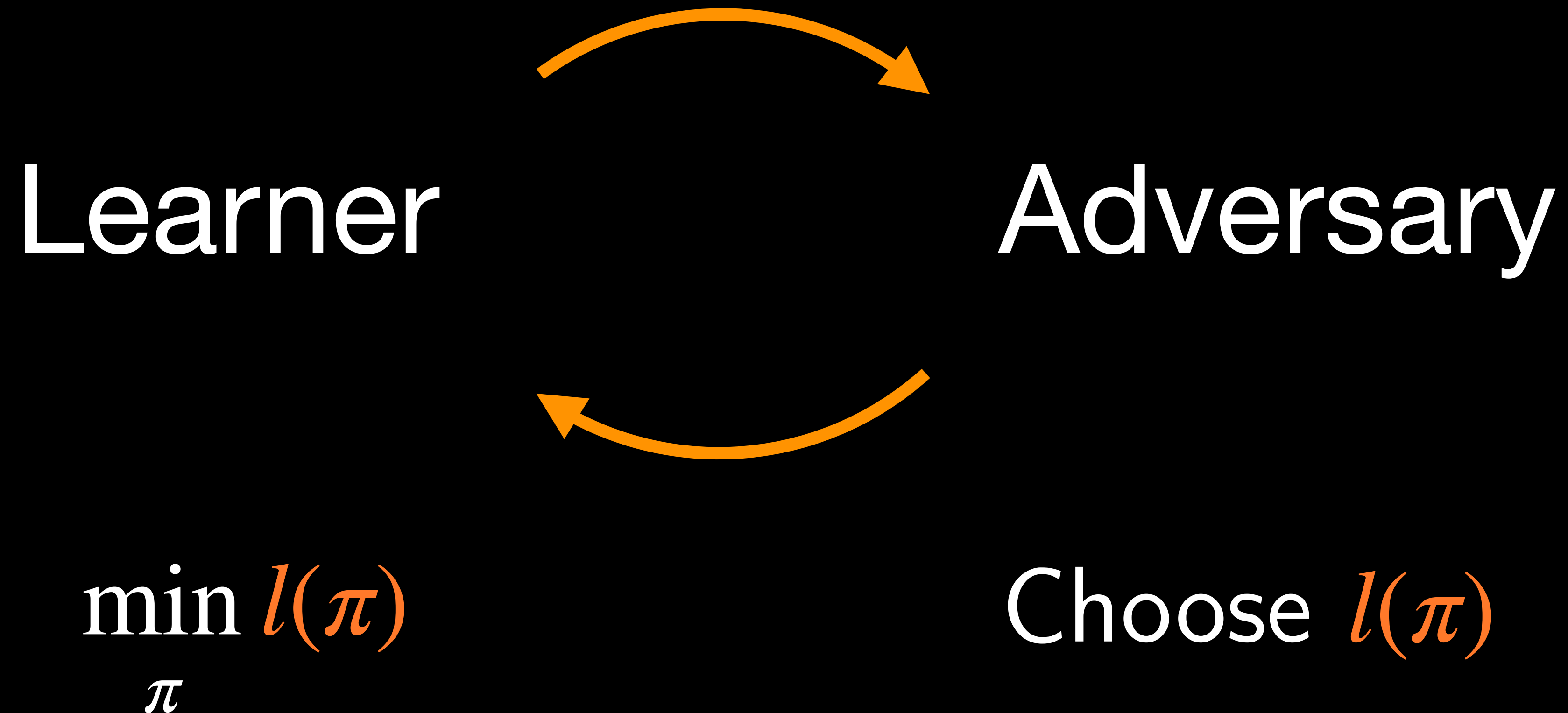
Don't know the test distribution upfront



$$\min_{\pi} l(\pi)$$

$$l(\pi) = \mathbb{E}_{x,y \sim P(x,y)} \ell(y, \pi(x))$$

Interactive Learning



Interactive Learning

Learner

Adversary

Initialize policy

π_1 [policy]

Chooses loss

$l_1(\cdot)$ [loss]

Update policy

π_2

Chooses loss

$l_2(\cdot)$

⋮

⋮

What is the best that I can do in such an adversarial setting?



Prediction
with
Expert Advice

Expert
2

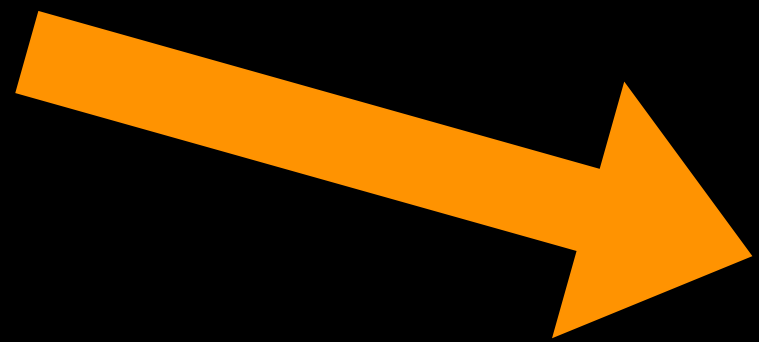
Expert 1



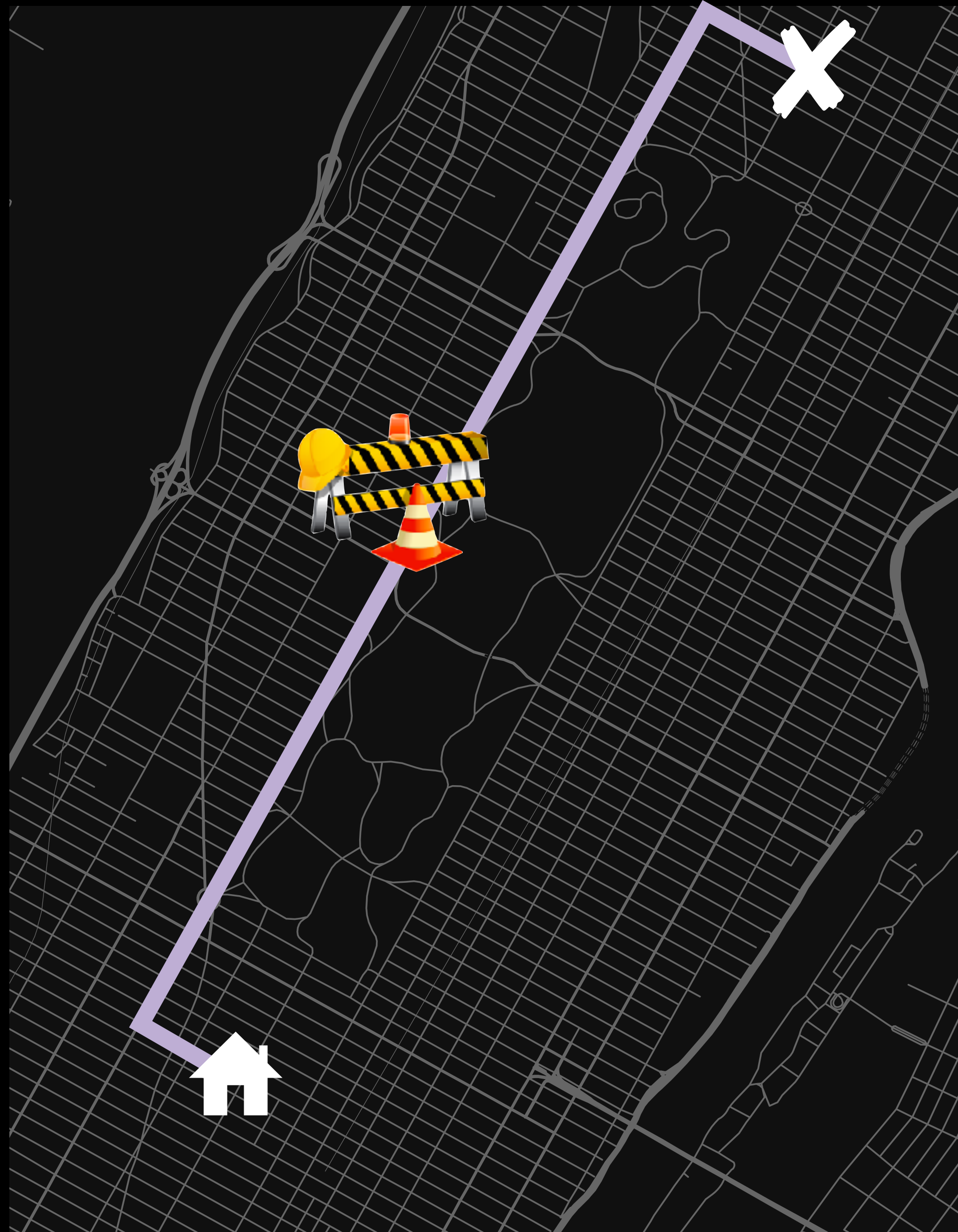
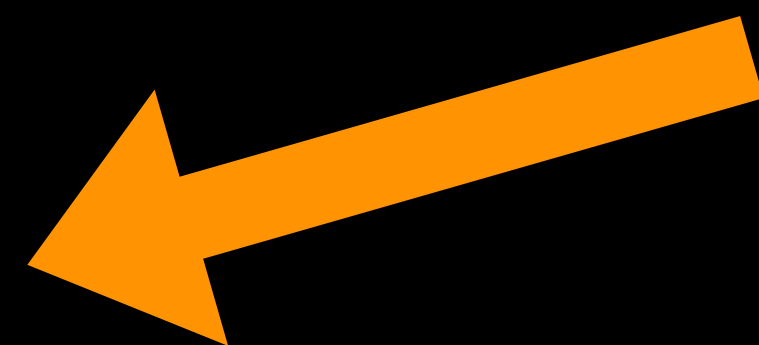
Expert
3



Expert 2



Loss = 1.0



Let's formalize!



$$\text{Regret} = \sum_{t=1}^T l_t(\pi_t) - \min_{\pi^*} \sum_{t=1}^t l_t(\pi^*)$$

(Learner)

(Best in
hindsight)

How do we design algorithms that are no-regret?

$$\text{Regret} = \sum_{t=1}^T l_t(\pi_t) - \min_{\pi^*} \sum_{t=1}^t l_t(\pi^*)$$



FOLLOW THE LEADER!



At every round t , choose the **best expert in hindsight**

$$\pi_t = \arg \min_{\pi} \sum_{i=1}^{t-1} l_i(\pi)$$

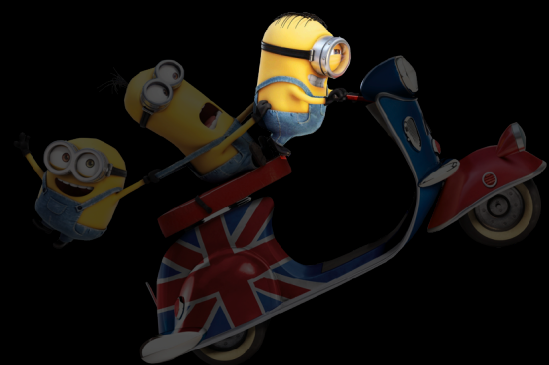
(lowest total loss)

$$\sum l_t$$

$$l_1$$

Expert 1

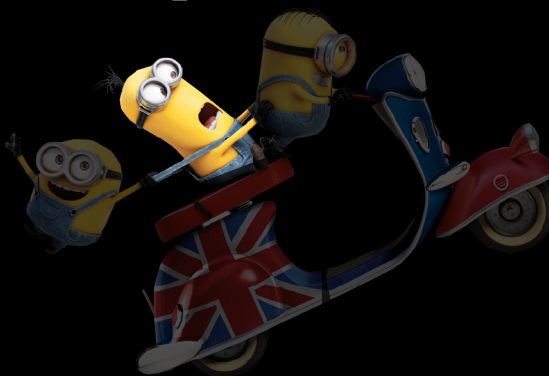
--



1.0

Expert 2

--



0.2

Expert 3

--



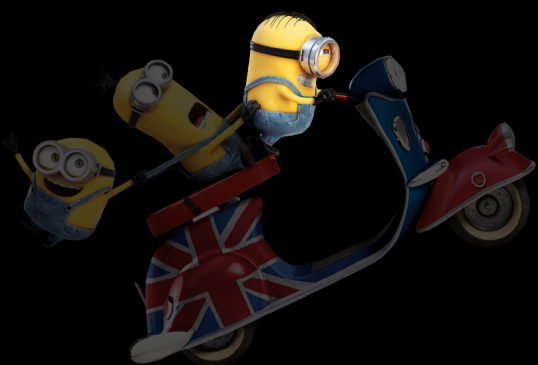
0.5

Avg. Regret: --

$\sum l_t$ l_1 l_2

Expert 1

1.0



1.0

0.5

Expert 2

0.2

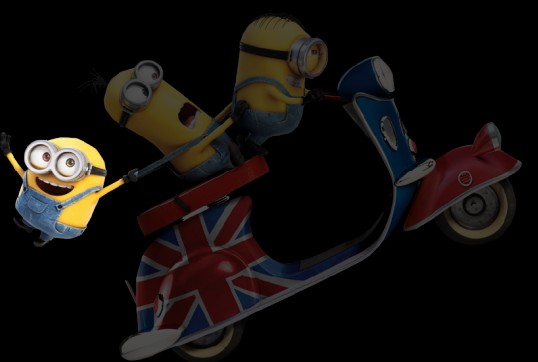


0.2

0.5

Expert 3

0.5



0.5

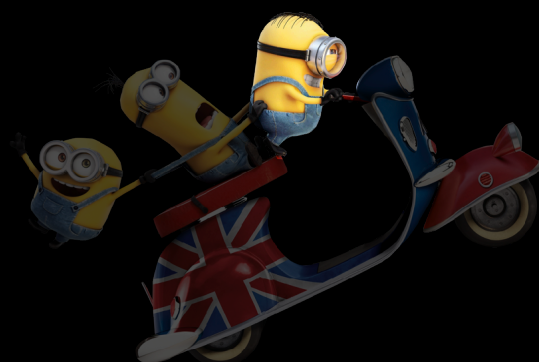
0.2

Avg. Regret: **0.80**

$\sum l_t$ l_1 l_2 l_3

1.5

Expert 1



1.0

0.5

0.5

0.7

Expert 2



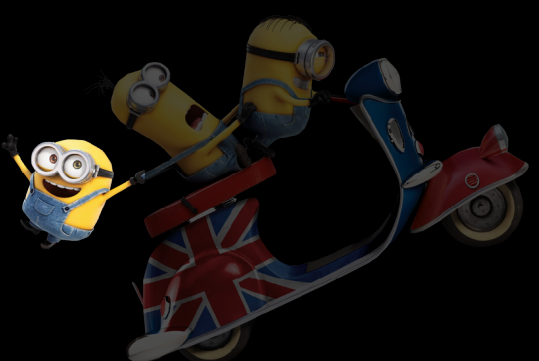
0.2

0.5

1.0

0.7

Expert 3



0.5

0.2

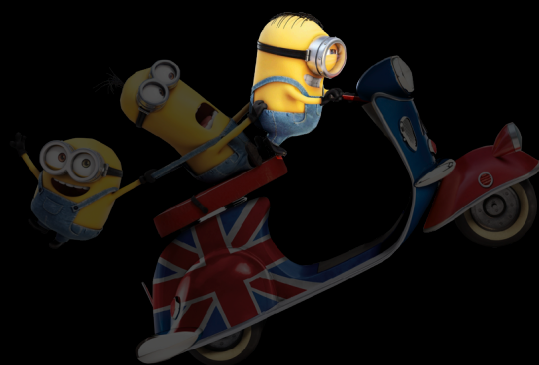
0.2

Avg. Regret: **0.40**

$\sum l_t$ l_1 l_2 l_3 l_4

Expert 1

2.0



1.0

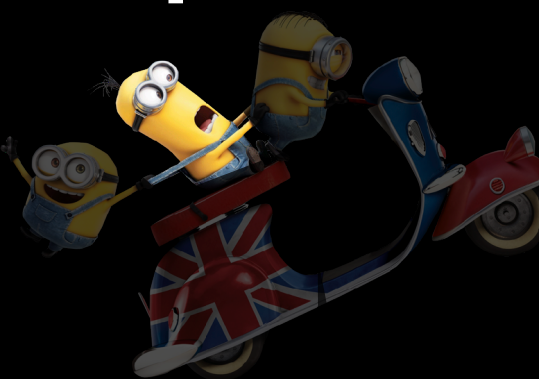
0.5

0.5

1.0

Expert 2

1.7



0.2

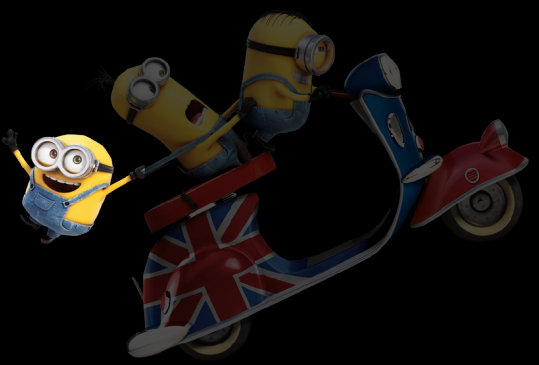
0.5

1.0

0.2

Expert 3

0.9



0.5

0.2

0.2

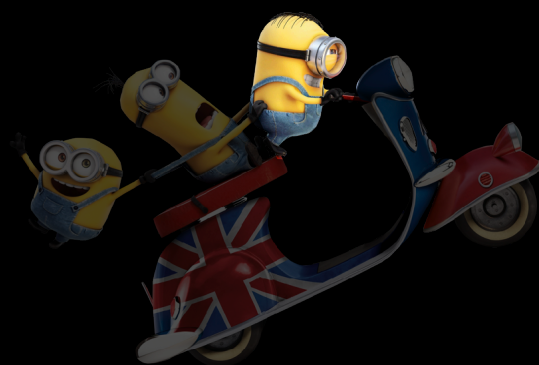
0.5

Avg. Regret: **0.53**

$\sum l_t$ l_1 l_2 l_3 l_4 l_5

Expert 1

3.0



1.0

0.5

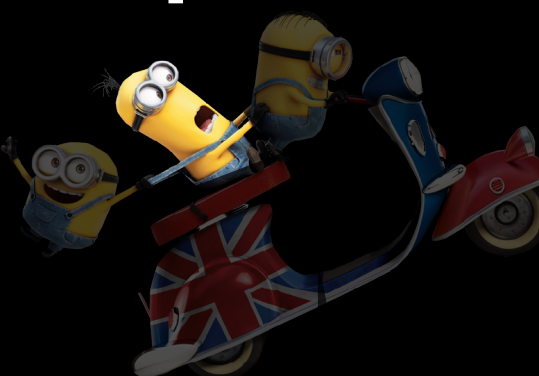
0.5

1.0

0.5

Expert 2

1.9



0.2

0.5

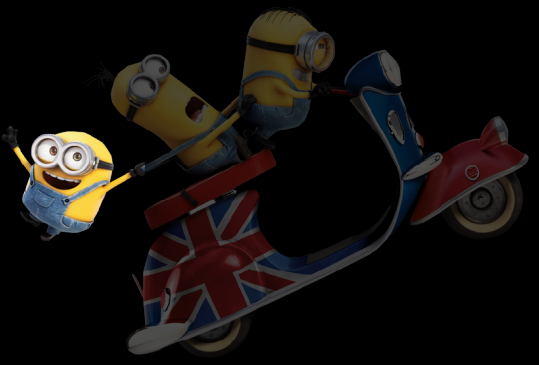
1.0

0.2

1.0

Expert 3

1.4



0.5

0.2

0.2

0.5

0.2

Avg. Regret: 0.40

$$\sum l_t$$

$$l_1$$

$$l_2$$

$$l_3$$

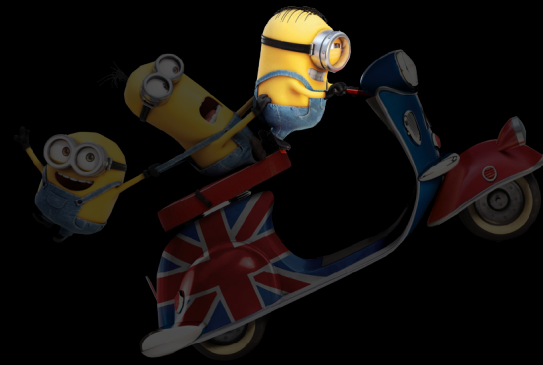
$$l_4$$

$$l_5$$

$$l_6$$

Expert 1

3.5



1.0

0.5

0.5

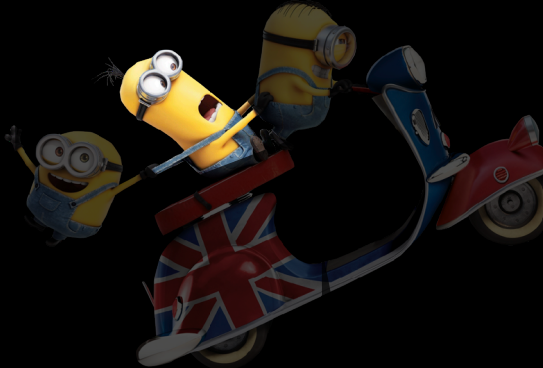
1.0

0.5

1.0

Expert 2

2.9



0.2

0.5

1.0

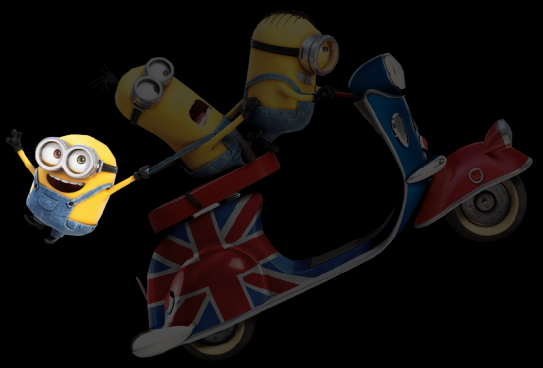
0.2

1.0

0.5

Expert 3

1.6



0.5

0.2

0.2

0.5

0.2

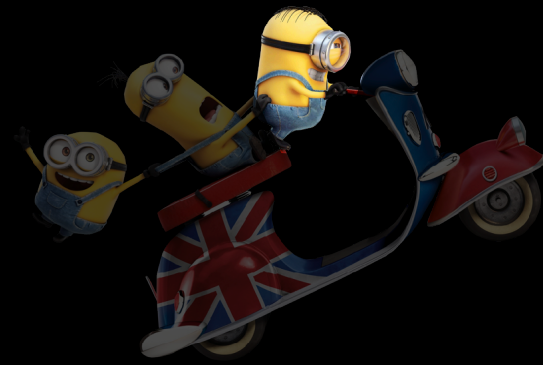
0.2

Avg. Regret: **0.32**

$\sum l_t$ l_1 l_2 l_3 l_4 l_5 l_6

Expert 1

4.5



1.0

0.5

0.5

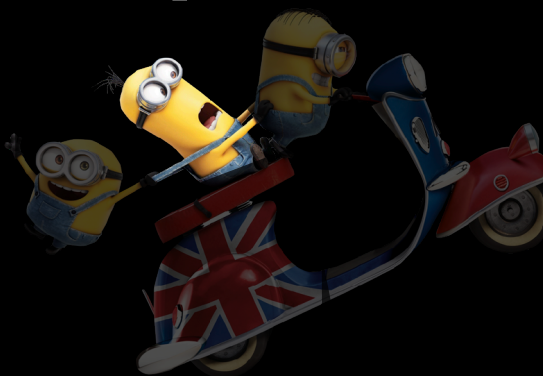
1.0

0.5

0.5

Expert 2

3.4



0.2

0.5

1.0

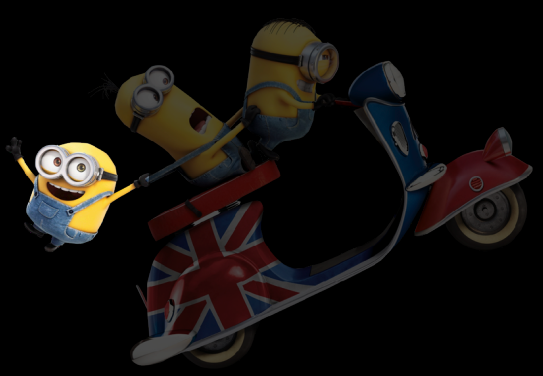
0.2

1.0

0.5

Expert 3

1.8



0.5

0.2

0.2

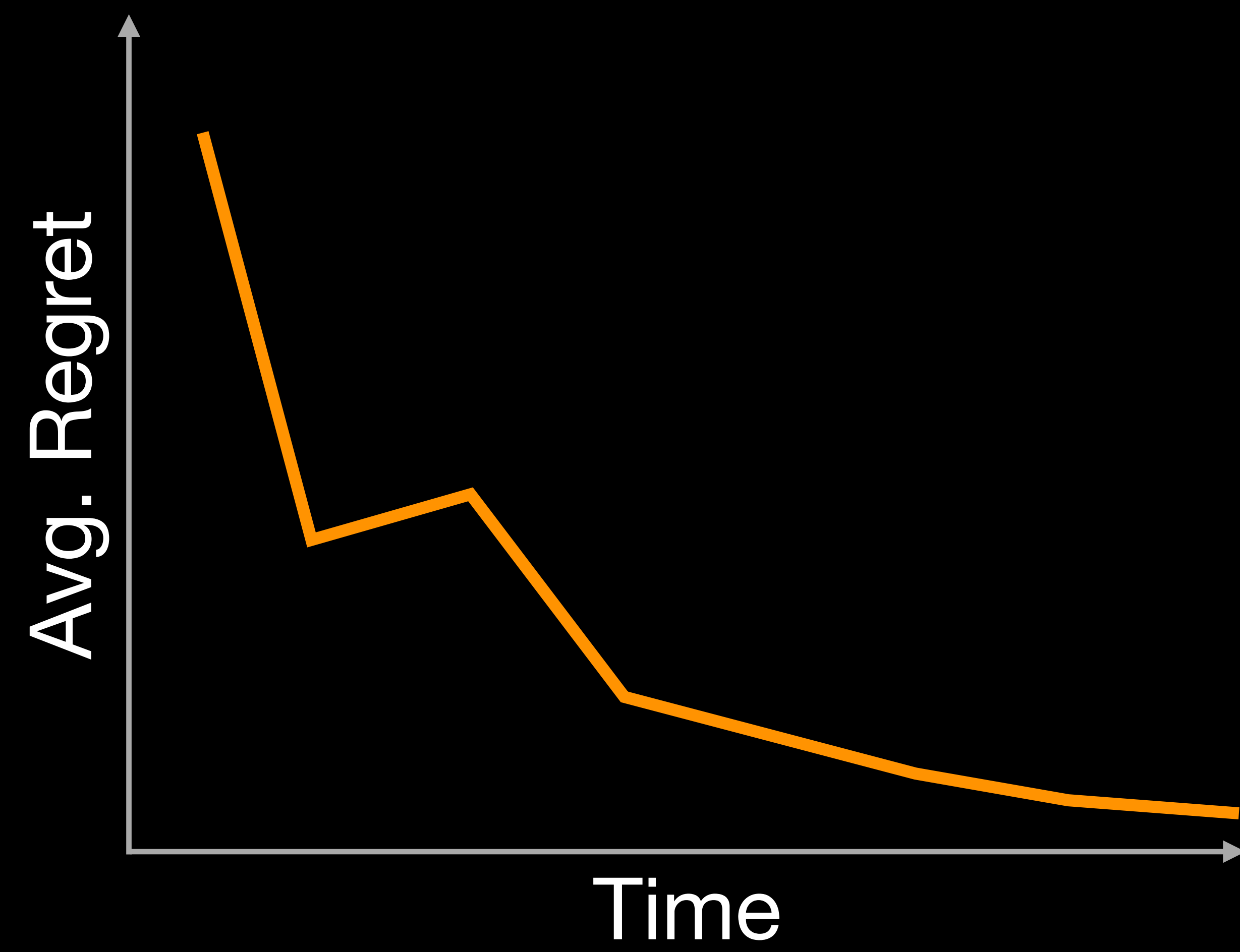
0.5

0.2

1.0

Avg. Regret: 0.26

FTL appears to be
no regret ...



Let's prove it!



Activity!



Think-Pair-Share!

Think (30 sec): Can you make FTL **high regret**?

Pair: Find a partner

Share (45 sec): Partners exchange ideas

FOLLOW THE LEADER!



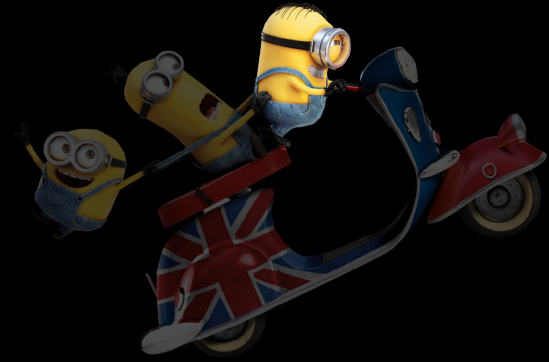
$$\pi_t = \arg \min_{\pi} \sum_{i=1}^{t-1} l_i(\pi)$$

(Hint: look at
the regret bound)

$$\sum l_t$$

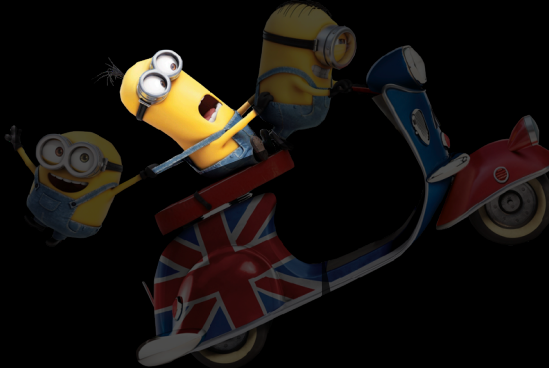
Expert 1

--

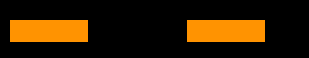


Expert 2

--



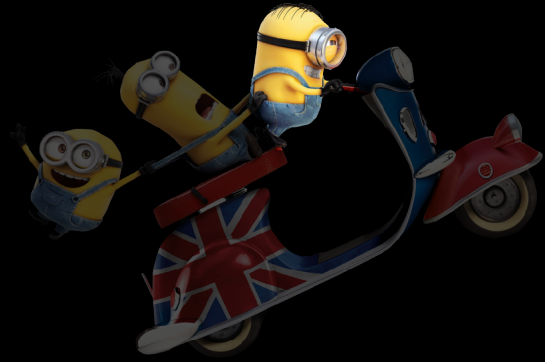
Avg. Regret: --



$$\sum l_t$$

Expert 1

--



$$l_1$$

1.0

Expert 2

--



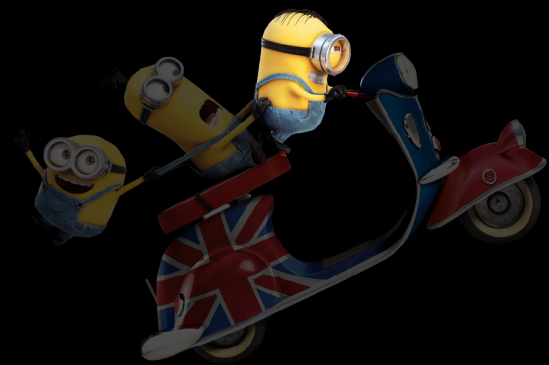
0.0

Avg. Regret: --

$\sum l_t$ l_1 l_2

Expert 1

1.0

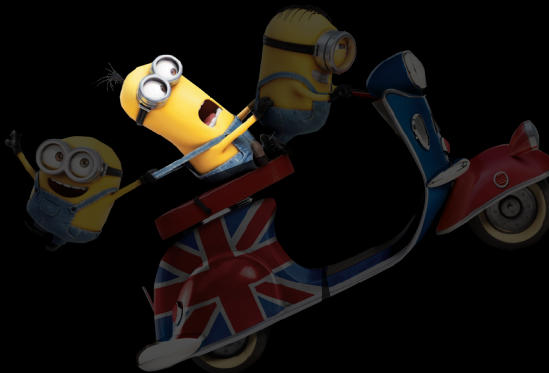


1.0

0.0

Expert 2

0.0



0.0

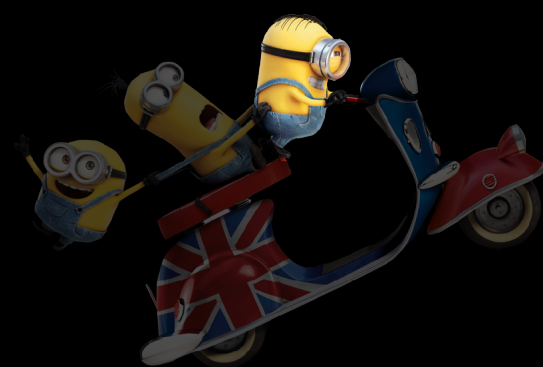
1.0

Avg. Regret: 1.00

$\sum l_t$ l_1 l_2 l_3

Expert 1

1.0



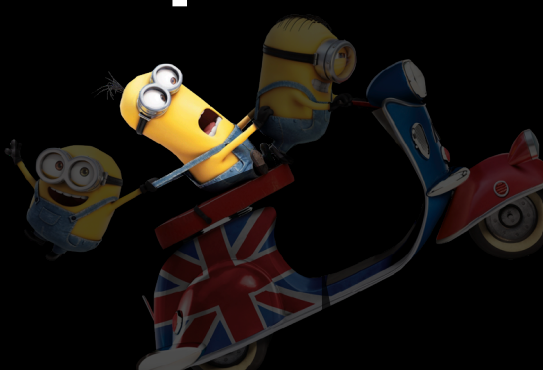
1.0

0.0

1.0

Expert 2

1.0



0.0

1.0

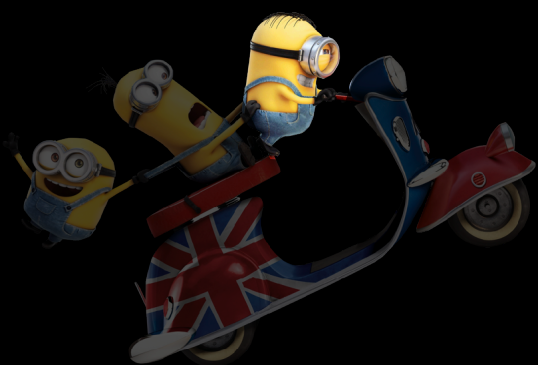
0.0

Avg. Regret: 0.50

$\sum l_t$ l_1 l_2 l_3 l_4

Expert 1

2.0



1.0

0.0

1.0

0.0

Expert 2

1.0



0.0

1.0

0.0

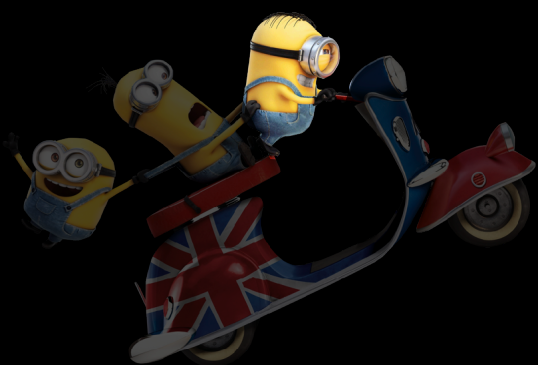
1.0

Avg. Regret: **0.67**

$\sum l_t$

2.0

Expert 1



1.0

0.0

1.0

0.0

1.0

 l_1 l_2 l_3 l_4 l_5

2.0

Expert 2



0.0

1.0

0.0

1.0

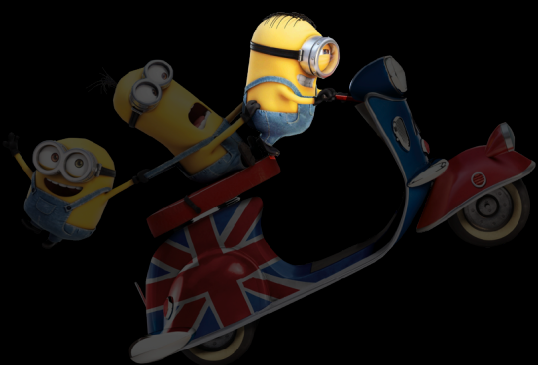
0.0

Avg. Regret: 0.50

$\sum l_t$ l_1 l_2 l_3 l_4 l_5 l_6

Expert 1

3.0



1.0

0.0

1.0

0.0

1.0

0.0

Expert 2

2.0



0.0

1.0

0.0

1.0

0.0

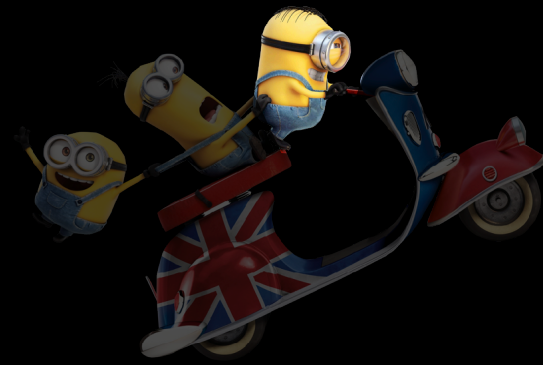
1.0

Avg. Regret: **0.60**

$\sum l_t$ l_1 l_2 l_3 l_4 l_5 l_6

Expert 1

3.0



1.0

0.0

1.0

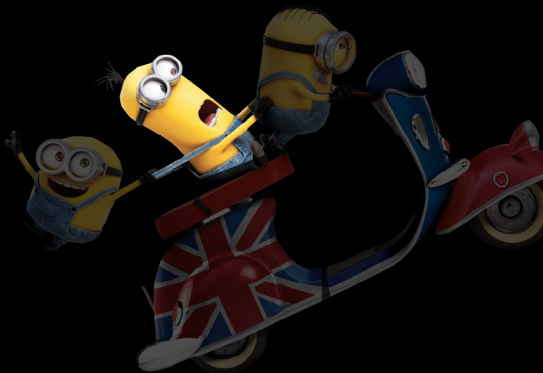
0.0

1.0

0.0

Expert 2

3.0



0.0

1.0

0.0

1.0

0.0

1.0

Predictions not stable \rightarrow High regret!

Avg. Regret: **0.50**

Cover's Impossibility Result

*“A powerful enough adversary
can drive the Regret of
any deterministic online algorithm
to $O(T)$
by anticipating its prediction
and setting maximal loss”*



How can we **curb the power** of the adversary?



π_t

$l_t(\pi_t)$



FOLLOW THE LEADER!

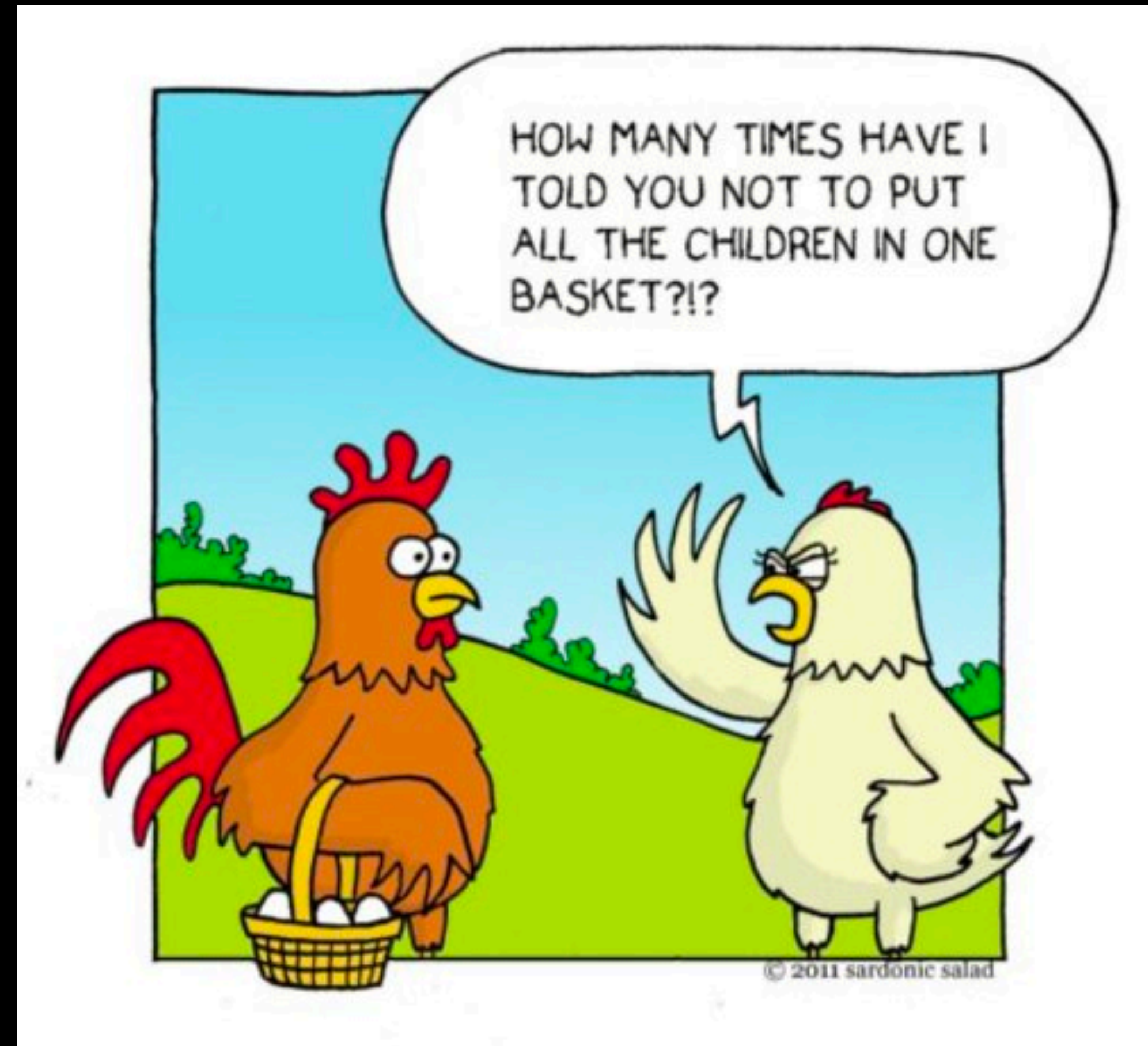


$$\pi_t = \arg \min_{\pi} \sum_{i=1}^{t-1} l_i(\pi)$$

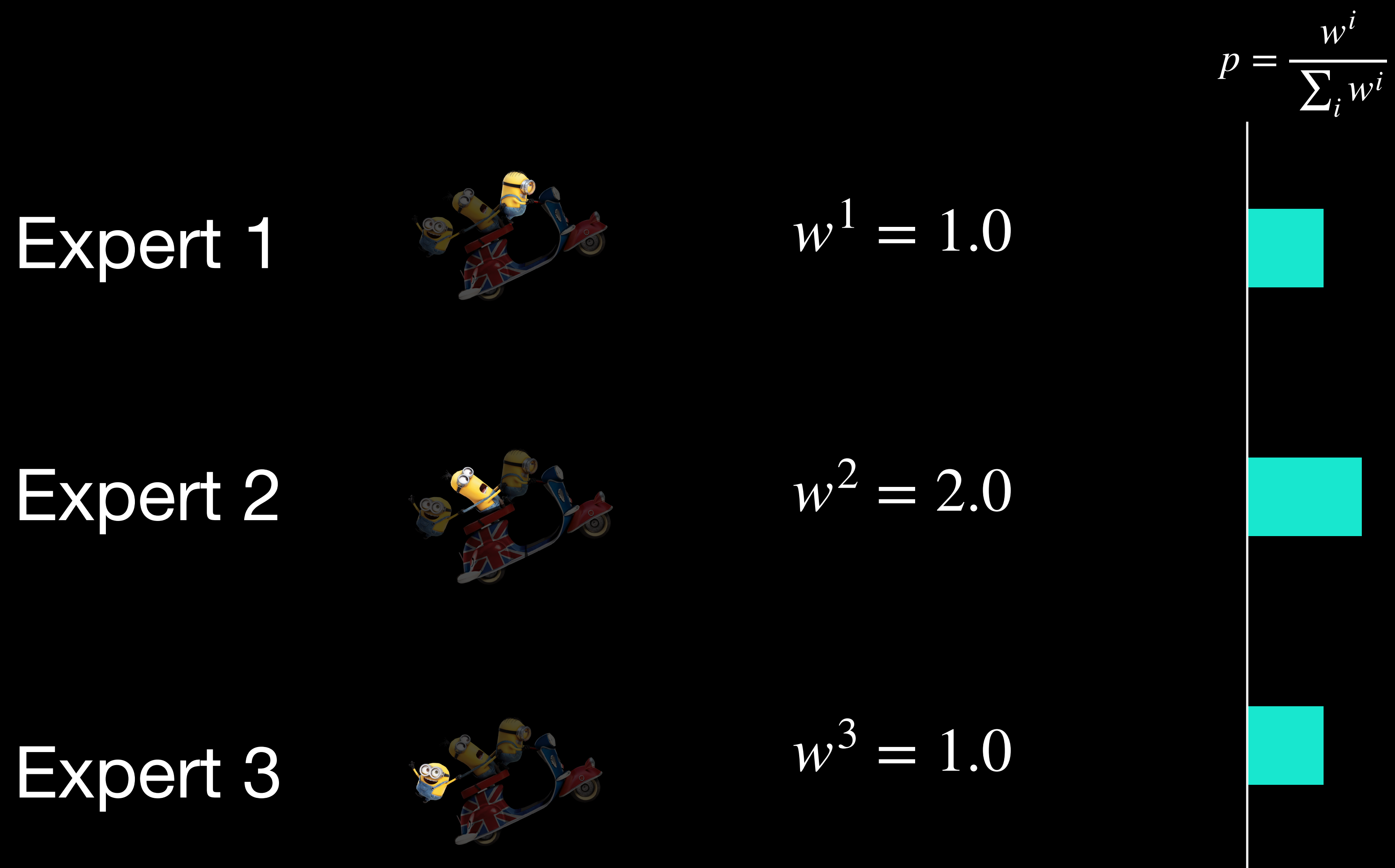
Adversary
breaks *any*
determinism



The virtue of hedging



Choose probability over experts



Let's formalize!



Let's apply FTL again (but on the space of weights)

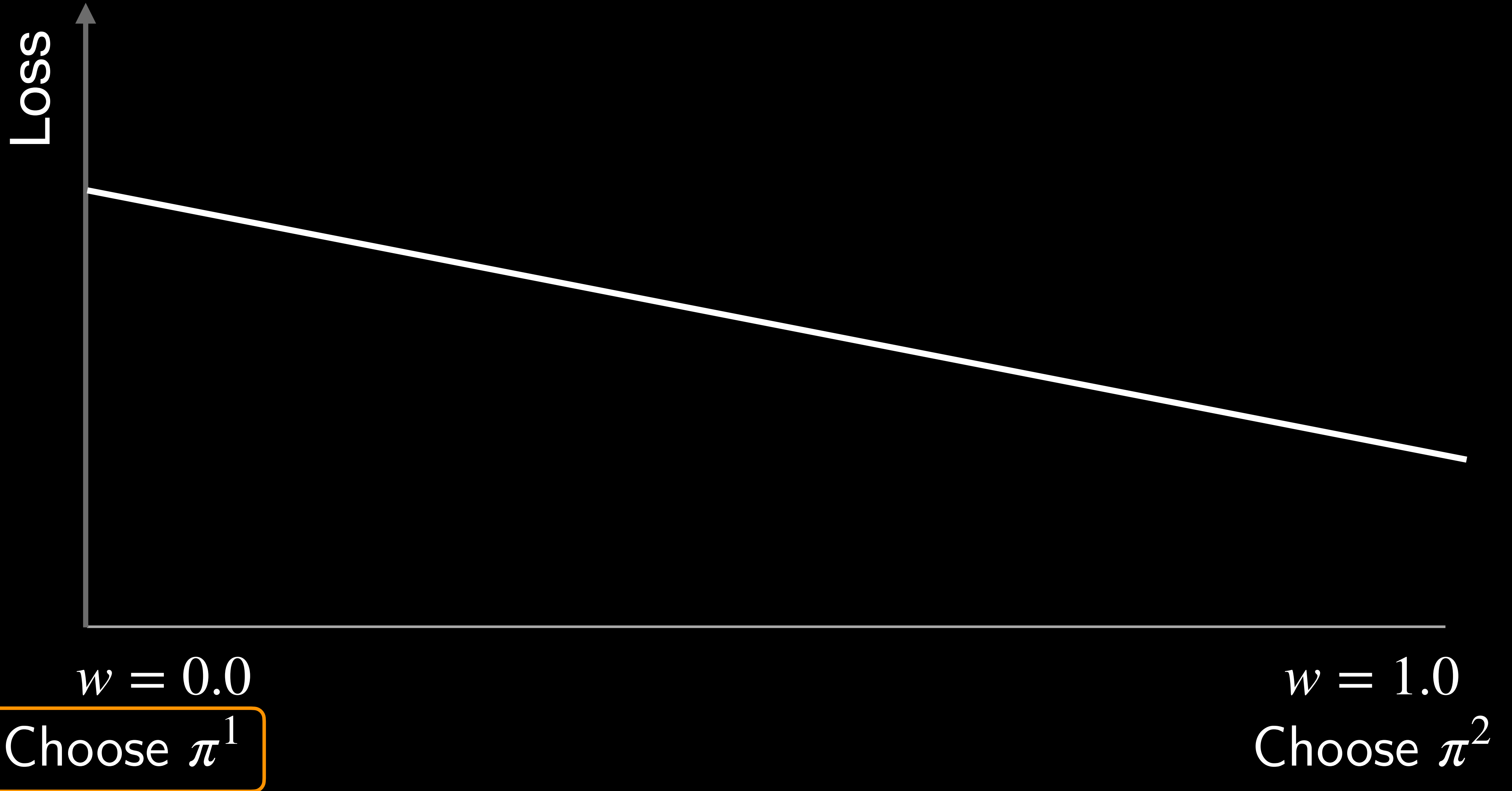
FOLLOW THE LEADER!



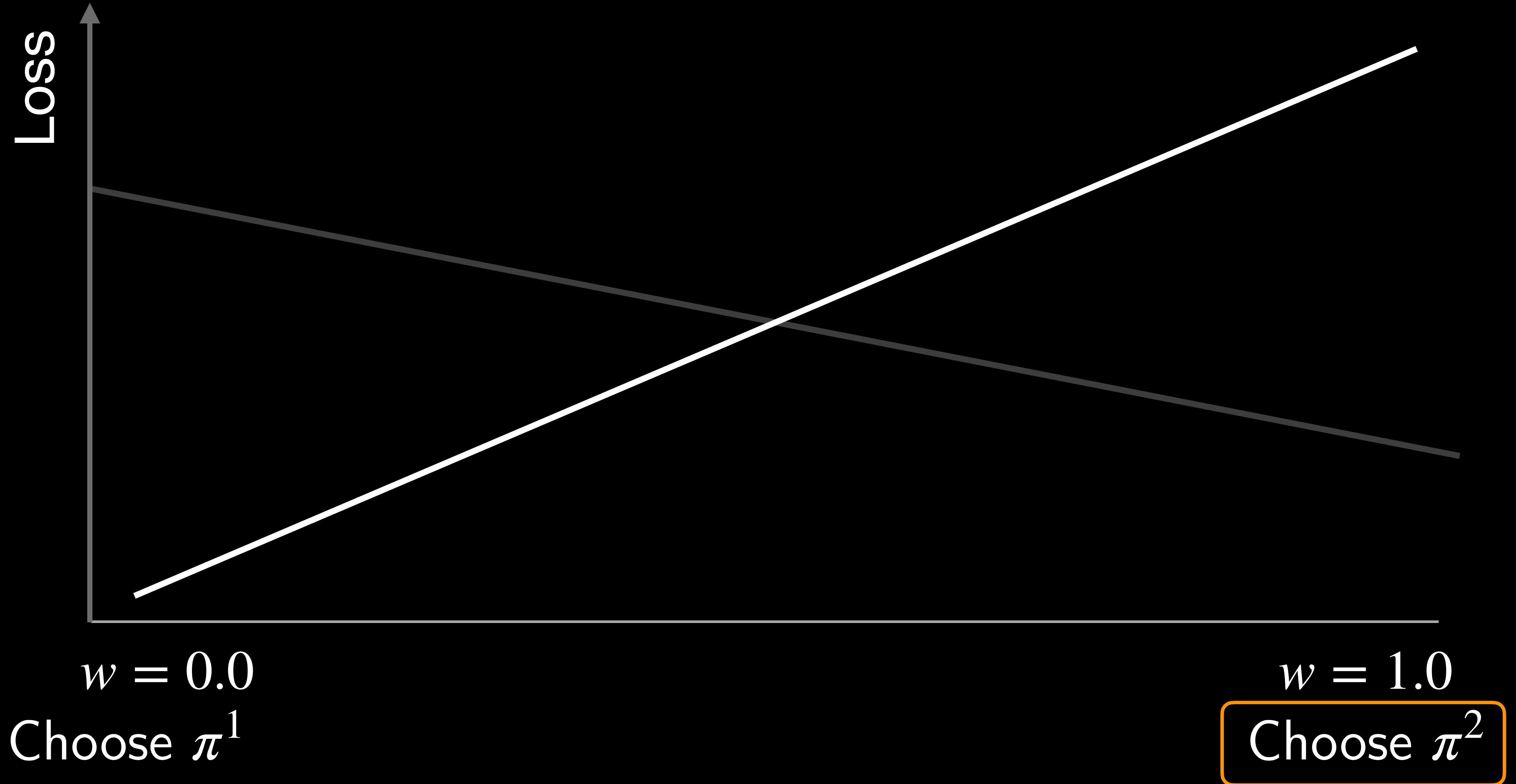
At every round t , choose the **best weights in hindsight**

$$w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w)$$

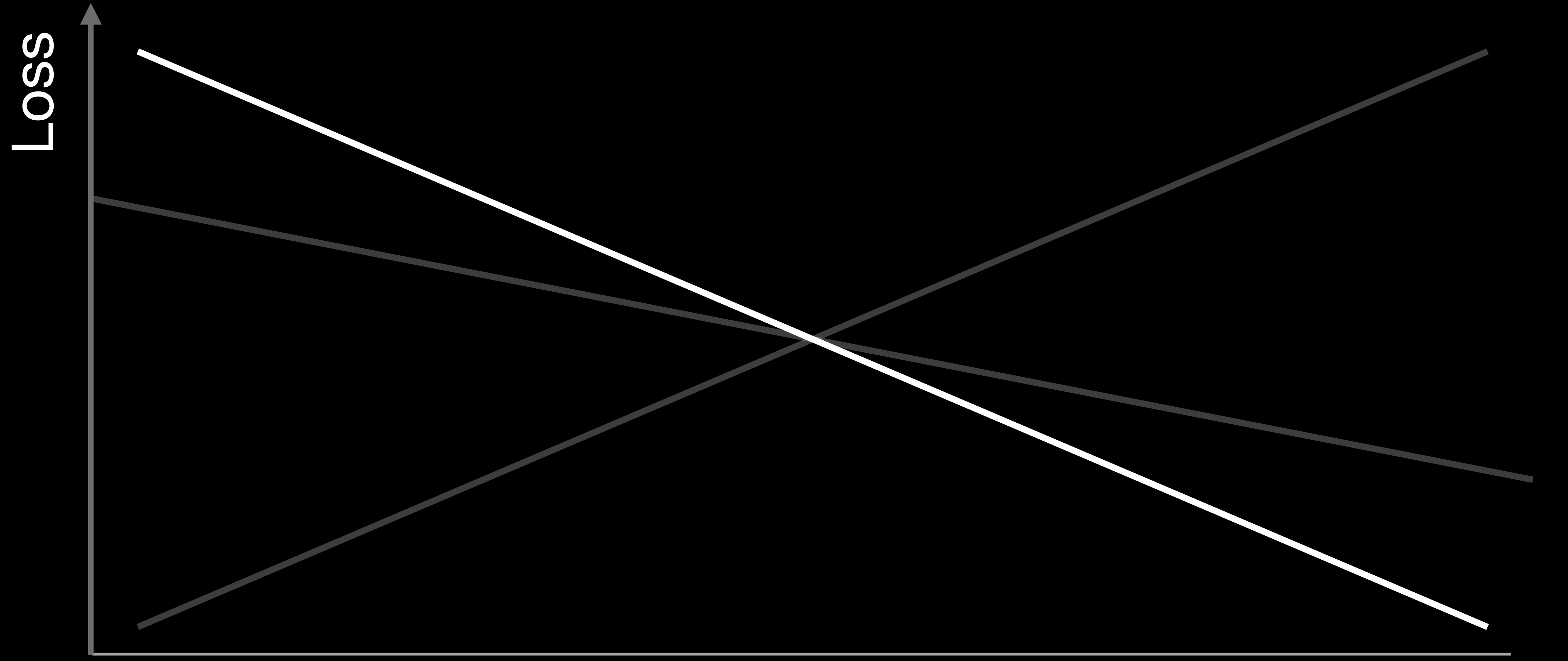
Loss = 0.75 Avg. Regret = 0.5



Loss = 1.0 Avg. Regret = 0.5



Loss = 1.0 Avg. Regret = 0.5



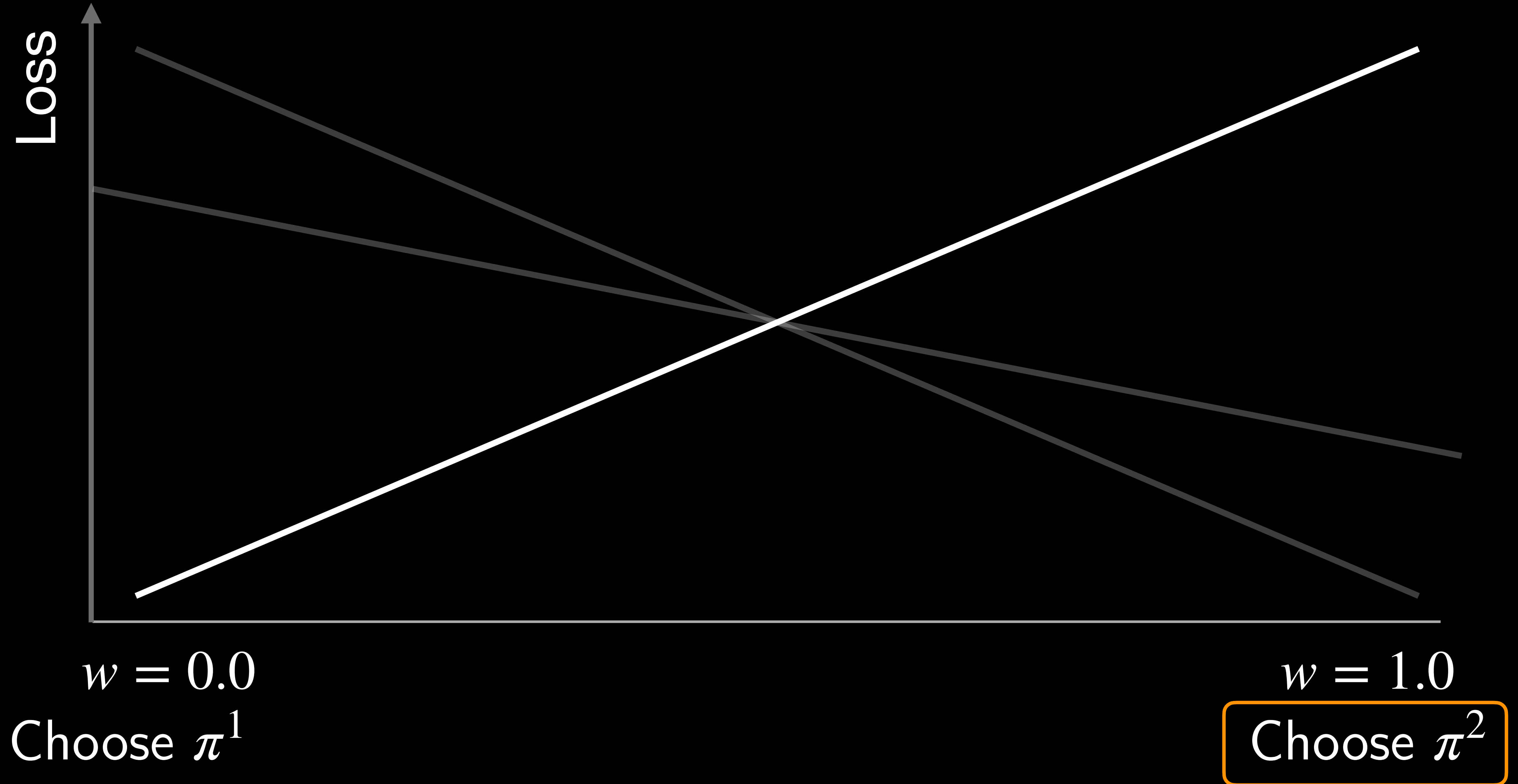
$w = 0.0$

Choose π^1

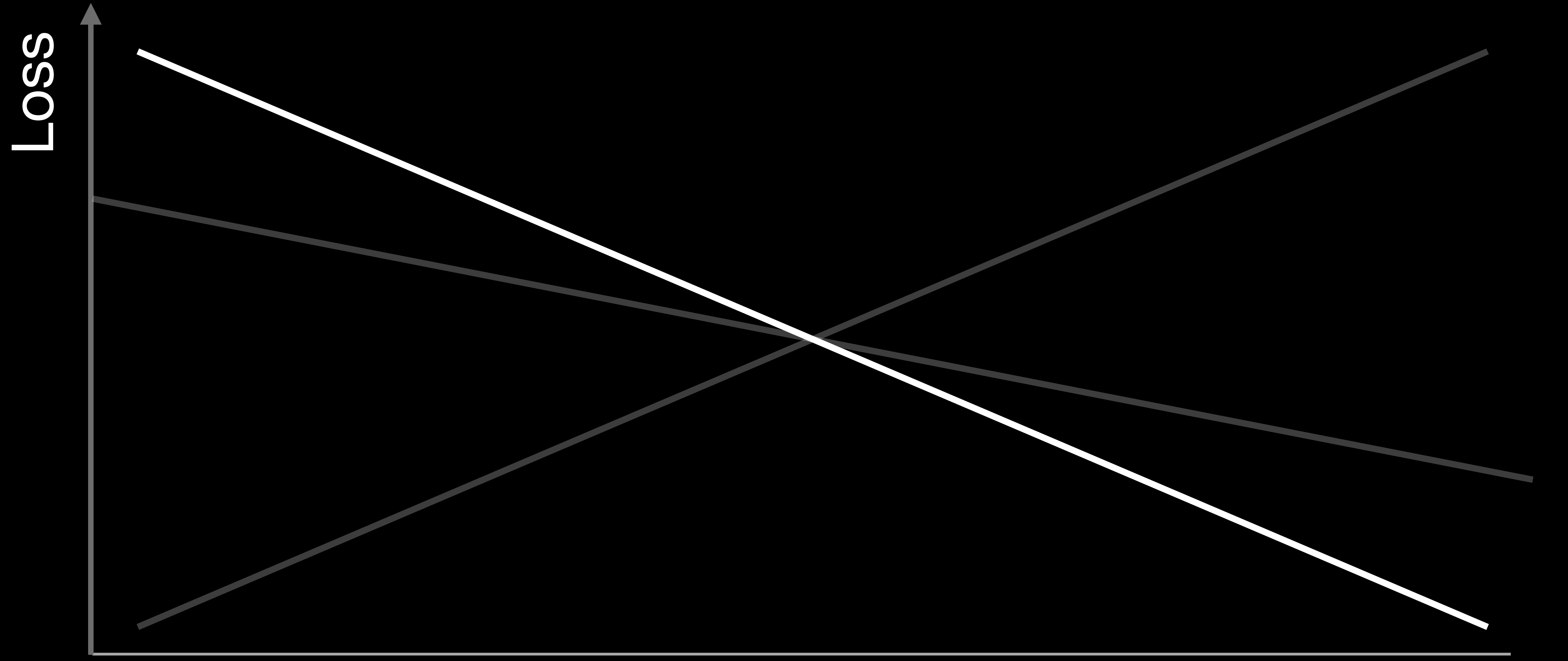
$w = 1.0$

Choose π^2

Loss = 1.0 Avg. Regret = 0.5



Loss = 1.0 Avg. Regret = 0.5



$w = 0.0$

Choose π^1

$w = 1.0$

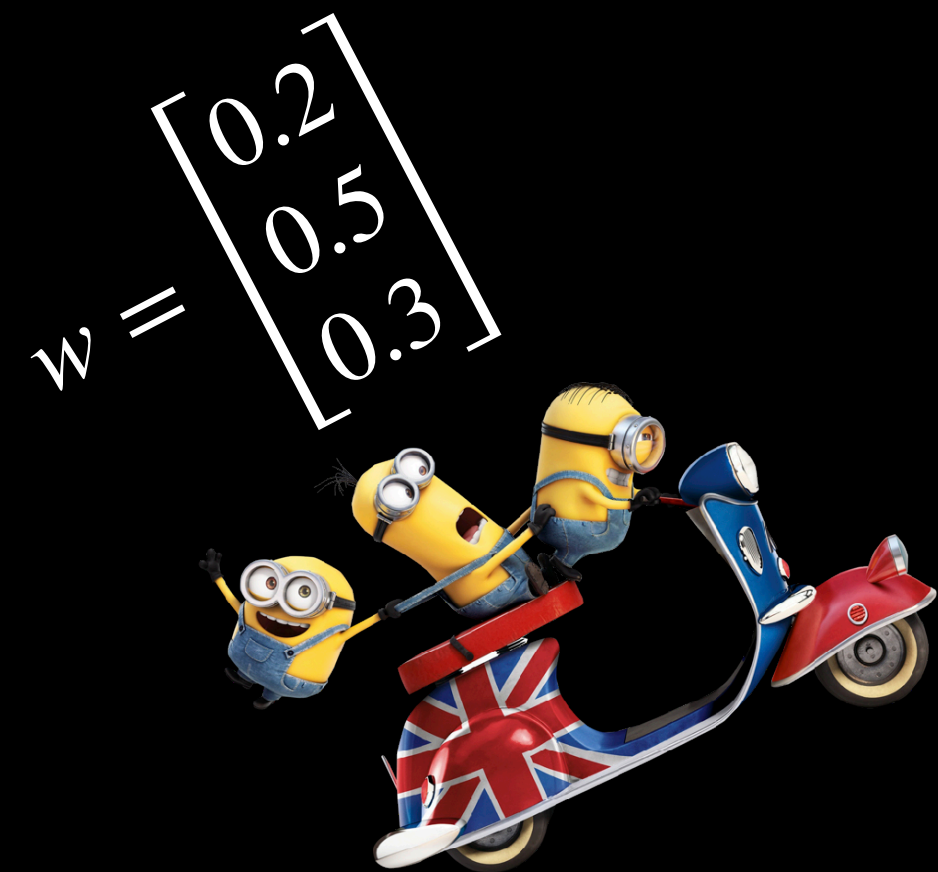
Choose π^2



Follow the leader
is too **aggressive** ...

Both in discrete and continuous settings!

Stability is the key problem!



w_t

$l_t(w_t)$



FOLLOW THE LEADER!



$$w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w)$$

Unstable predictions!



Be stable

Slowly change
predictions



Follow the Regularized Leader



$$w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w) + \eta_t R(w)$$

Strong regularization!

Regularization ensures $\|w_t - w_{t+1}\| \rightarrow 0$ as $t \rightarrow \infty$

Follow the Regularized Leader



$$w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w) + \eta_t R(w)$$

Strong regularization!

What are some choices for regularization?

GENERALIZED WEIGHTED

MAJORITY

Episode IV

A NEW HOPE

GENERALIZED WEIGHTED MAJORITY

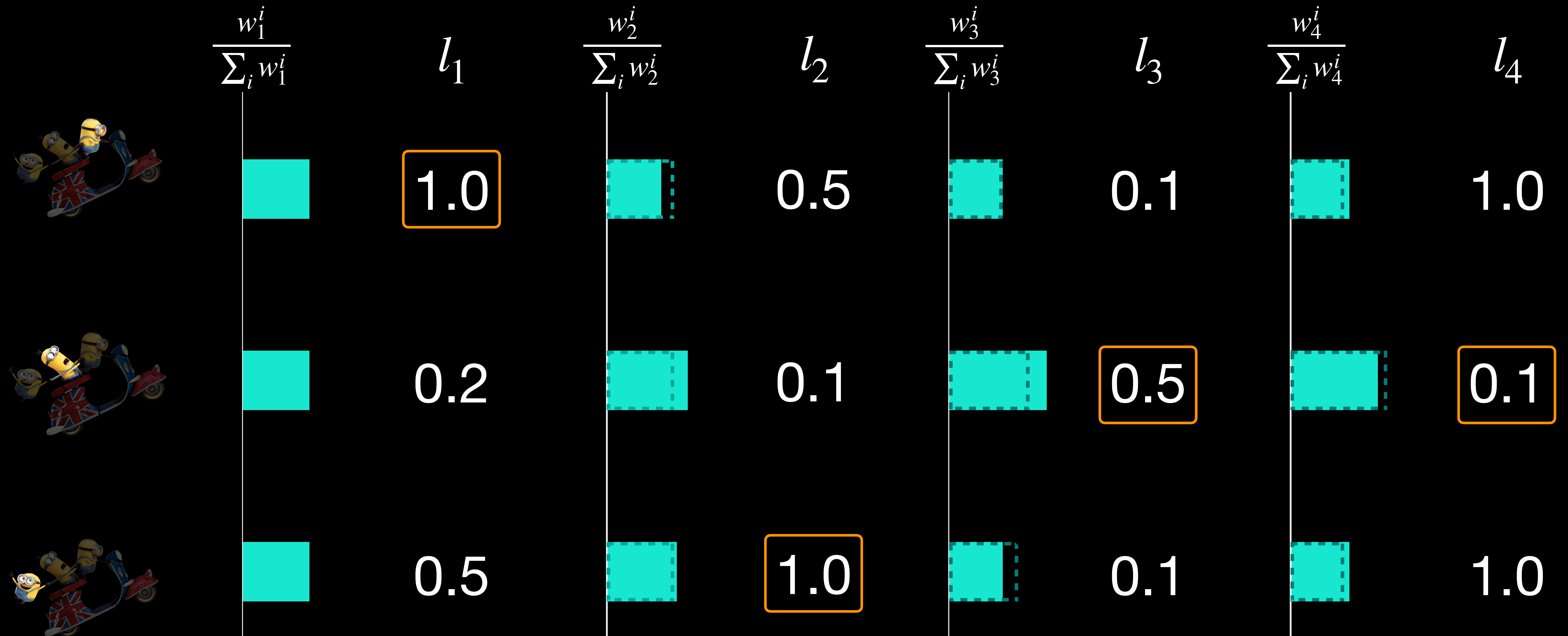
1. At $t=1$, set weight for expert i as $w_1^i = 1$

2. At time t , choose expert i with probability $\frac{w_t^i}{\sum_i w_t^i}$

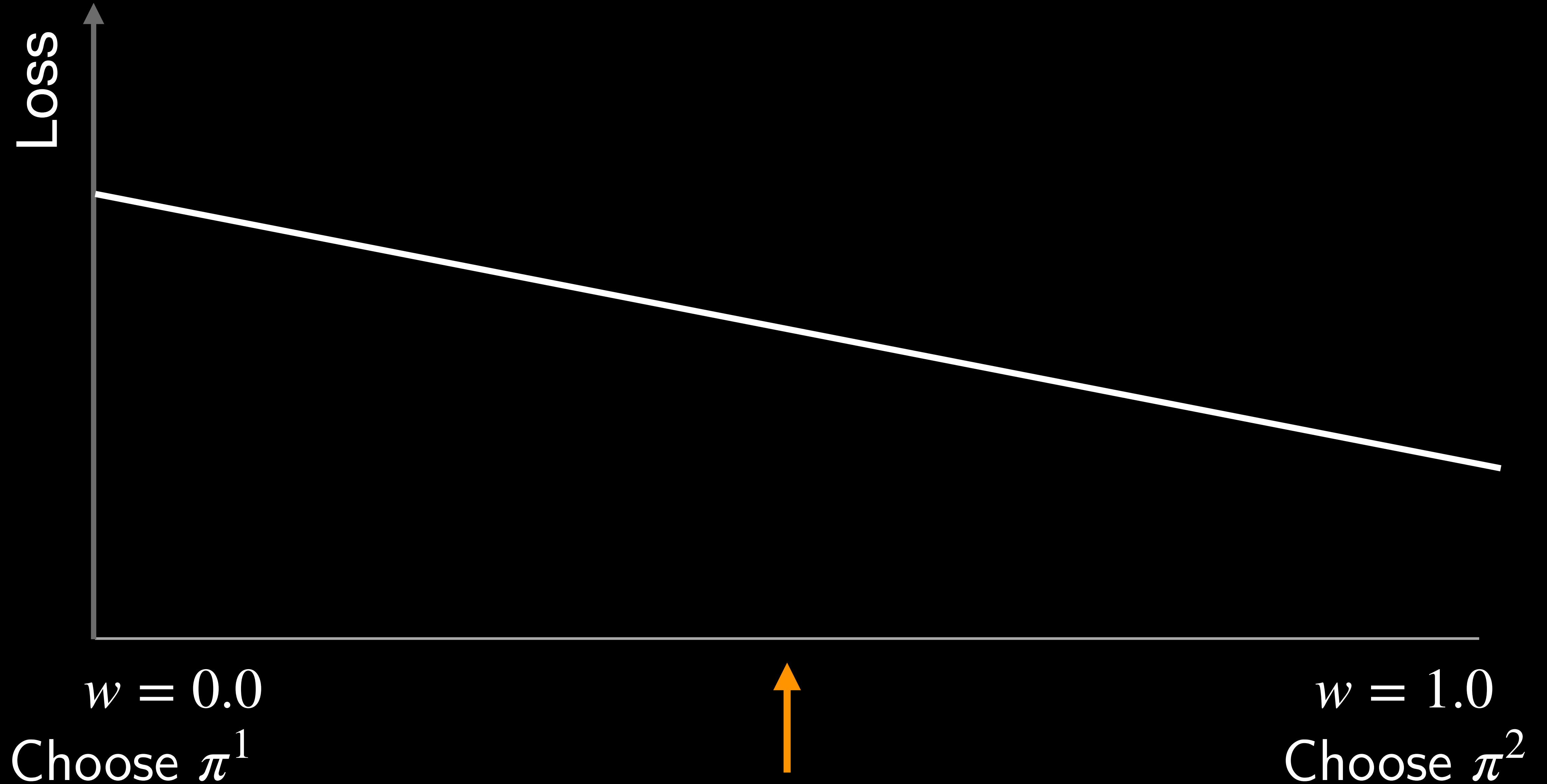
3. Update weight for expert i (Bump down if loss is high)

$$w_{t+1}^i = w_t^i \exp(-\eta l_t(\pi^i))$$

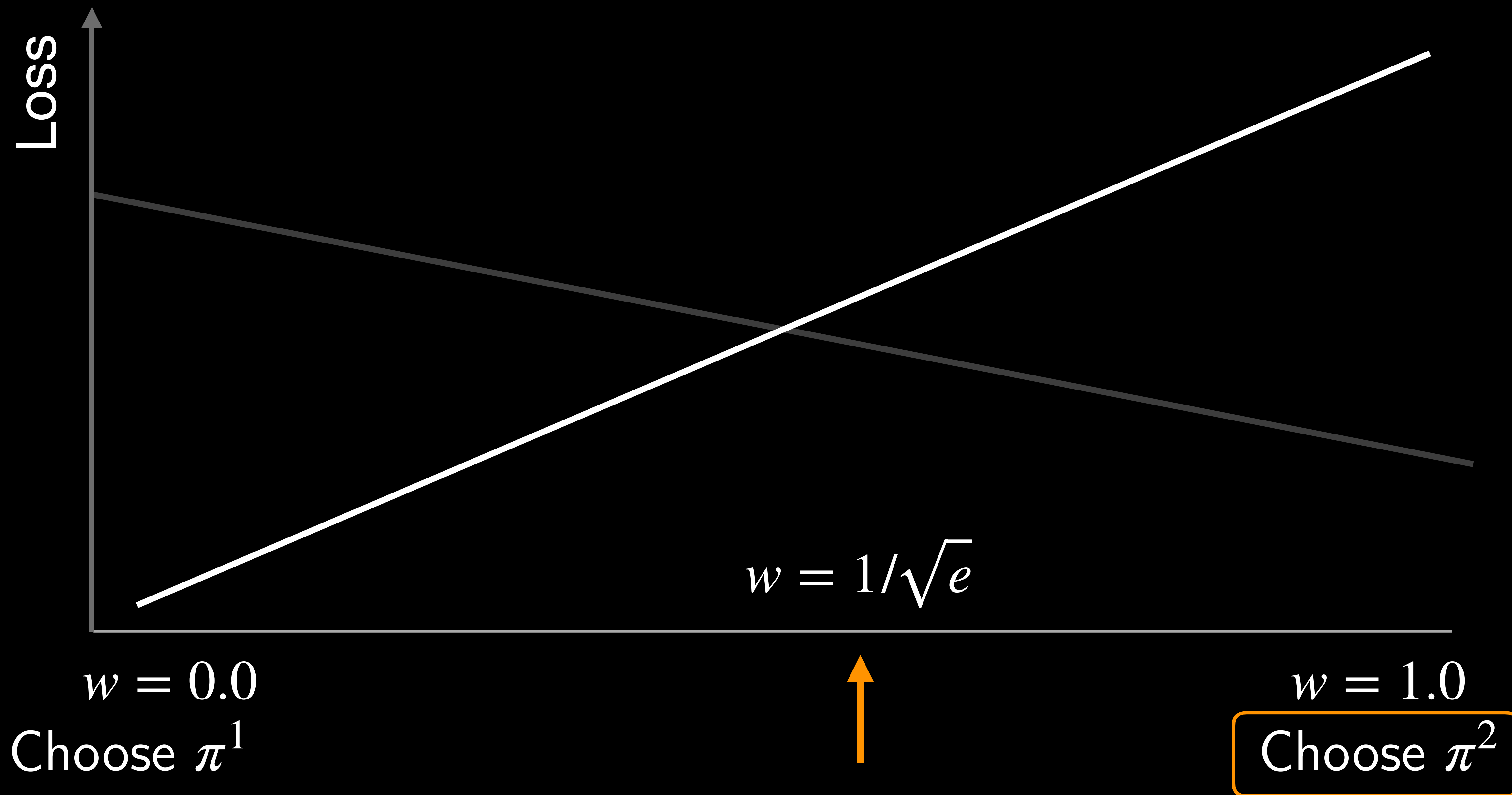
GENERALIZED WEIGHTED MAJORITY



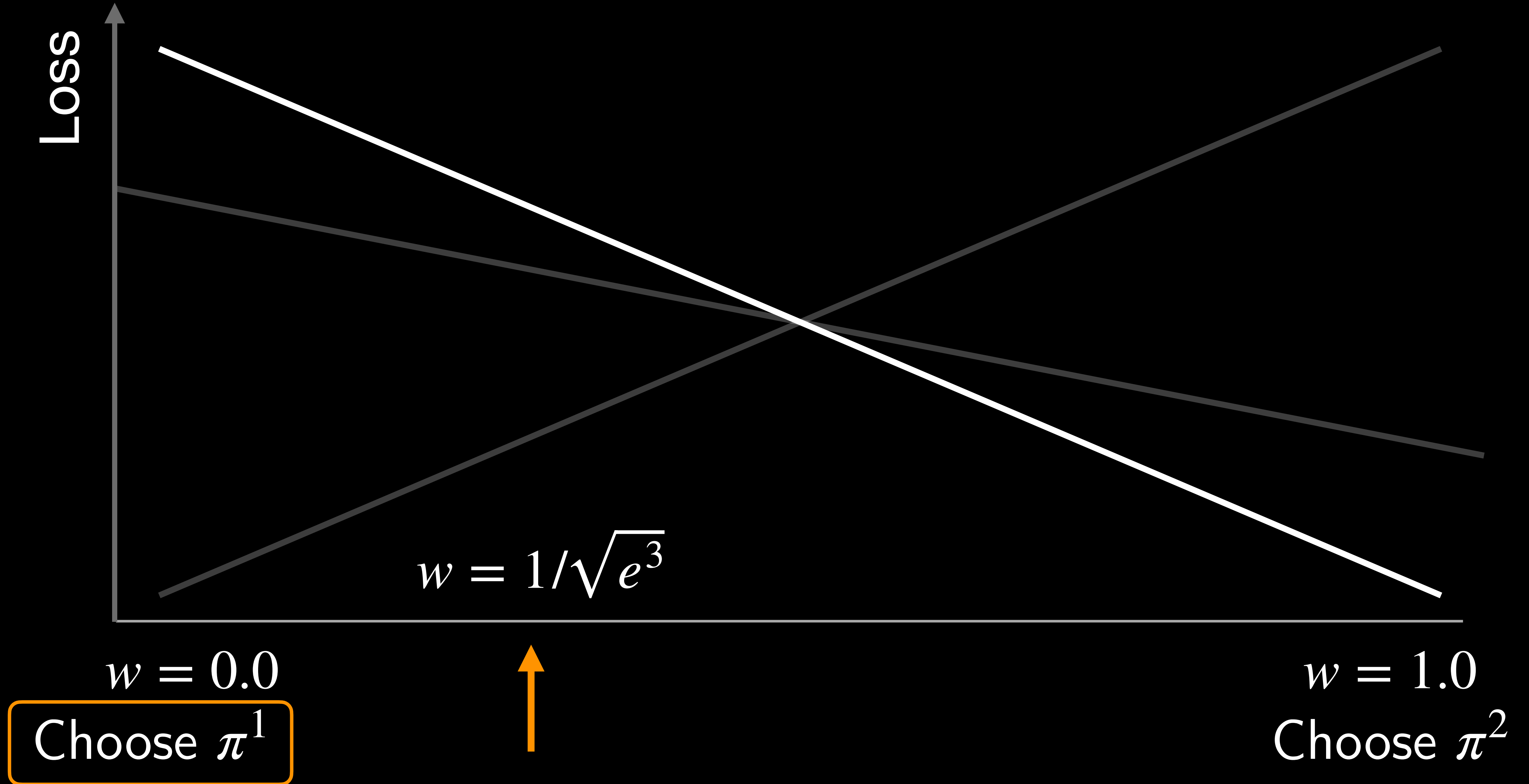
Loss = 0.5 Avg. Regret = 0.25



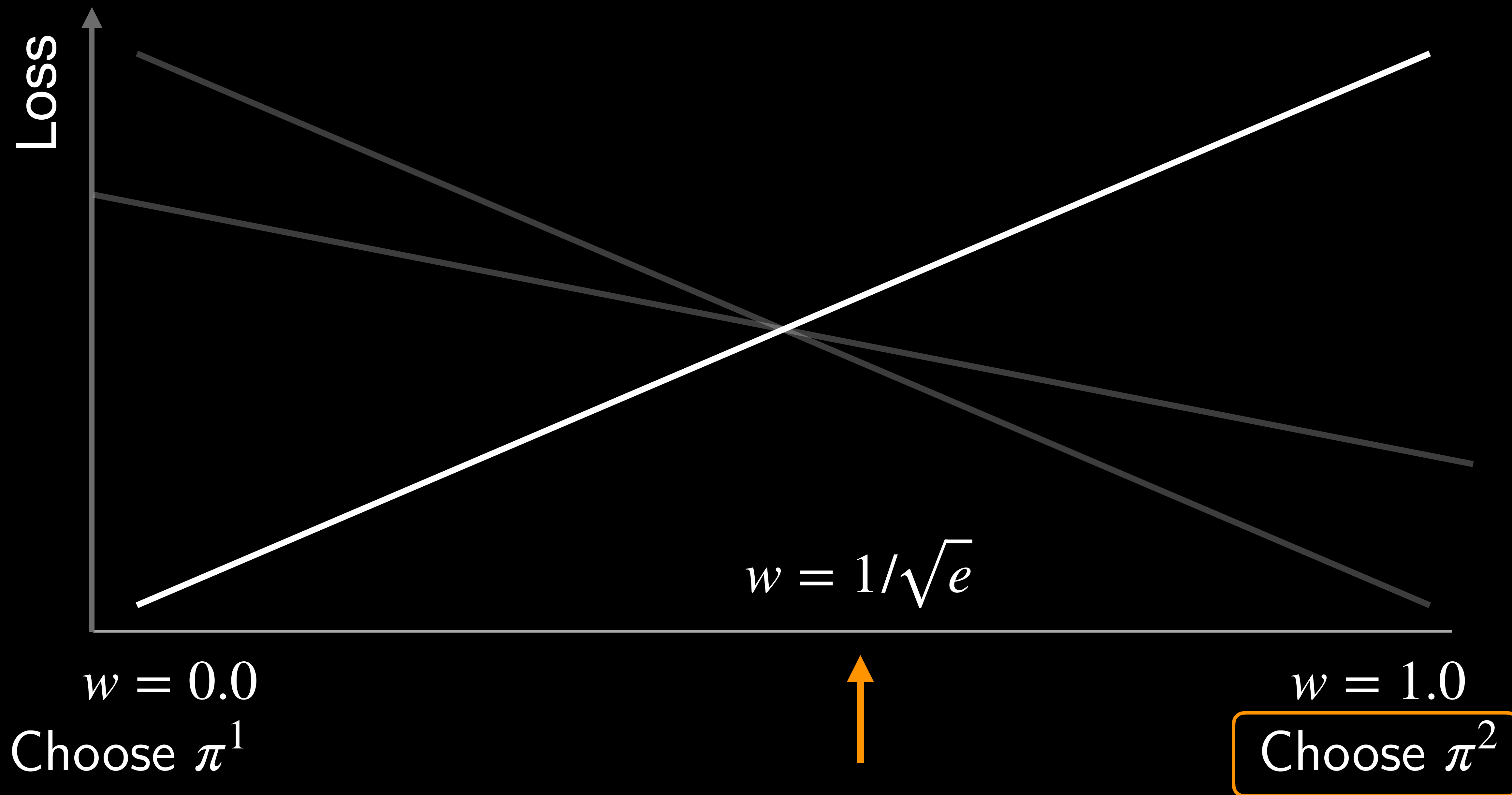
Loss = 0.6 Avg. Regret = 0.17



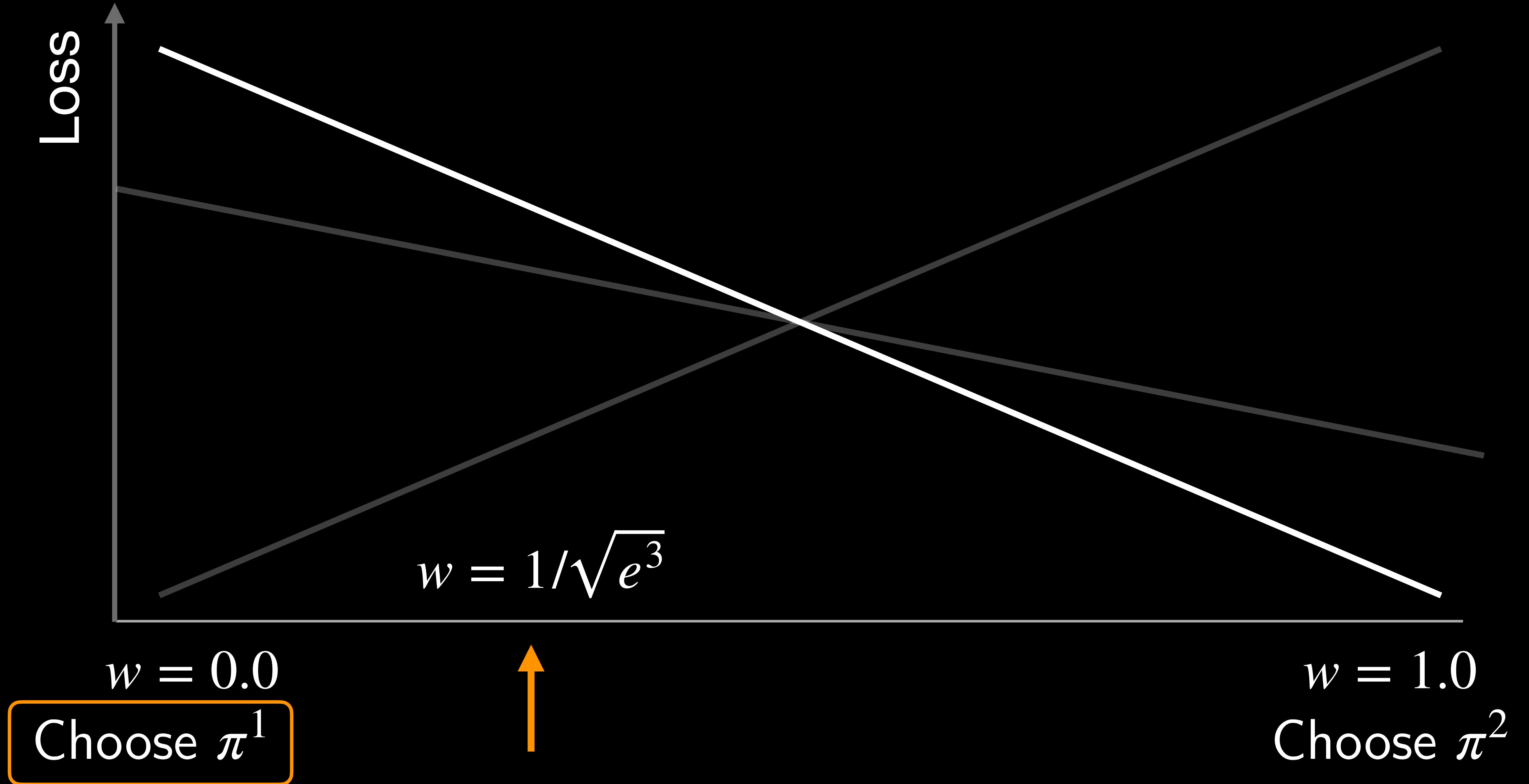
Loss = 0.78 Avg. Regret = 0.21



Loss = 0.6 Avg. Regret = 0.18



Loss = 0.78 Avg. Regret = 0.2



Linear
Programming

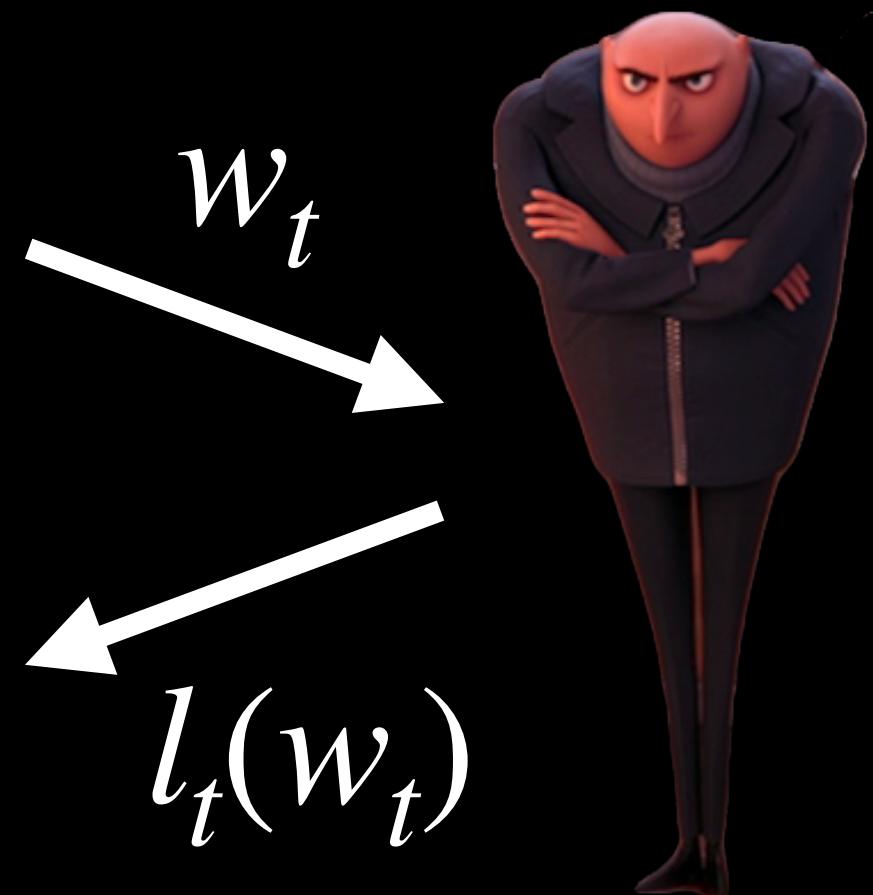
Boosting

Games

Soft-RL



$$w = \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix}$$


 w_t
 $l_t(w_t)$


$$w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w)$$

Regularization
 \Rightarrow No Regret!



$$w_t = \arg \min_w \sum_{i=1}^{t-1} l_i(w) + \eta_t R(w)$$

Unstable predictions!

