# Building Classification Models

# Two Approaches to Deadlines

Start 5 minutes before deadline

Good luck with that

Start 1 year before deadline

Maybe overkill

Neither approach is optimal

# Starting a Year in Advance

Probability of meeting the deadline

100%

Probability of getting other important work done

0%

# Starting Five Minutes in Advance

Probability of meeting the deadline

0%

Probability of getting other important work done

100%

# The Goldilocks Solution

## Work fast

Start very late and hope for the best

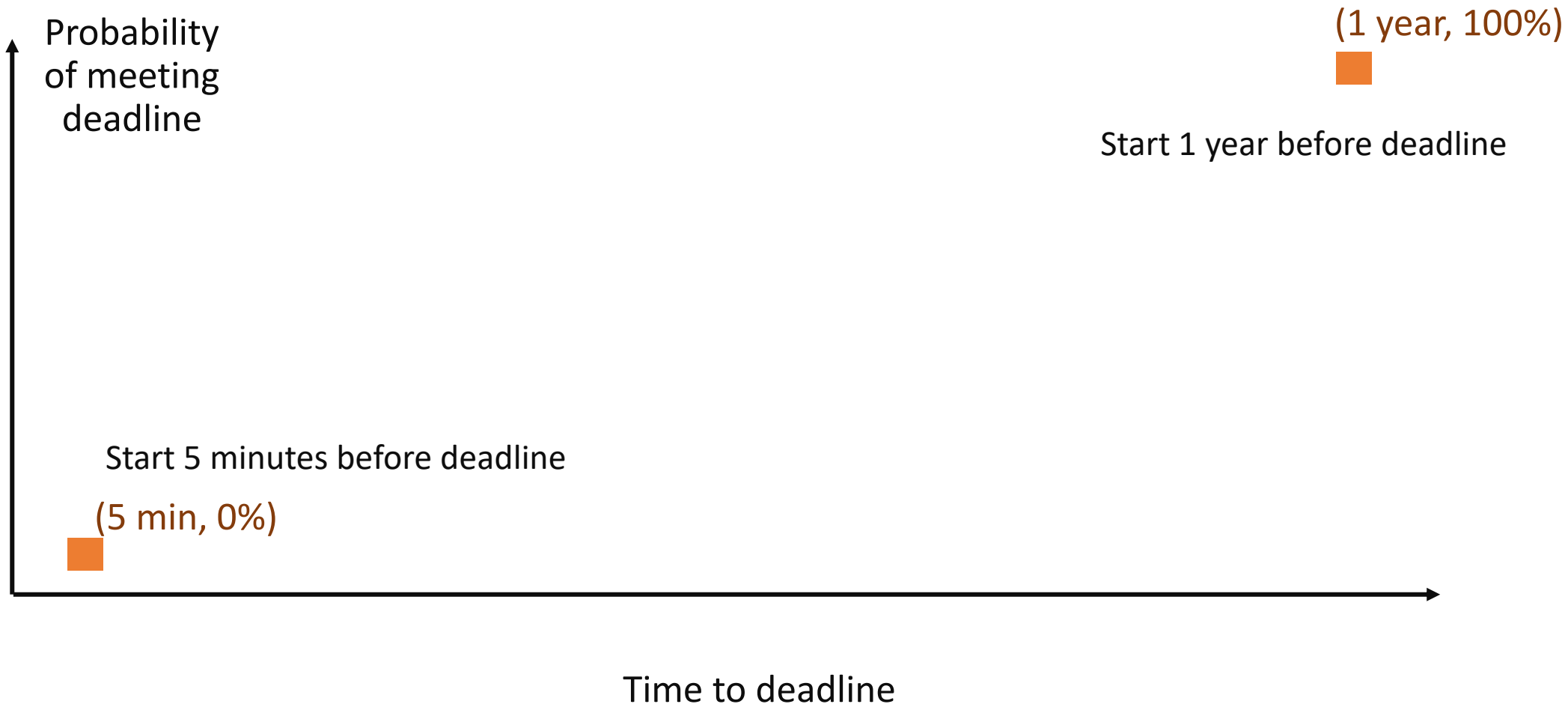
## Work smart

Start as late as possible To be sure to make it

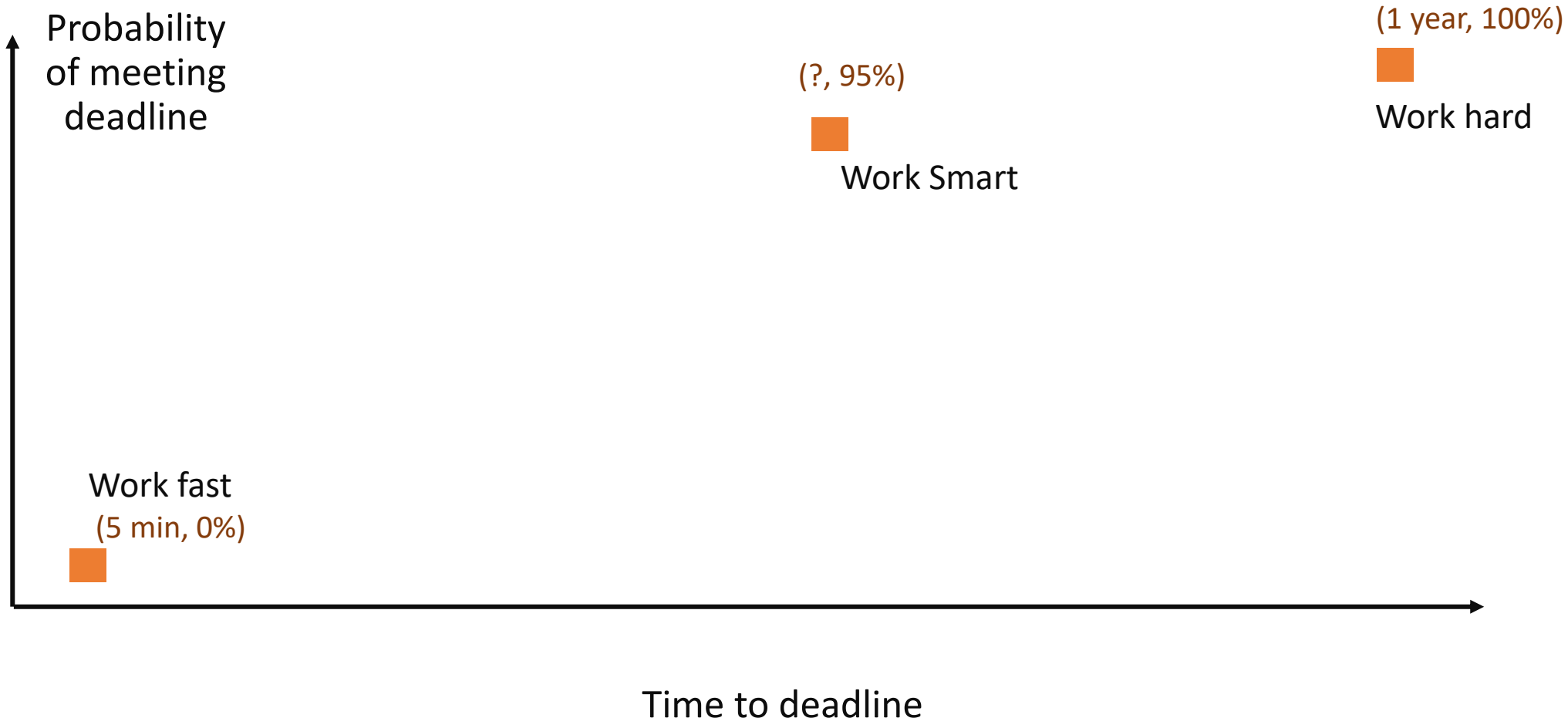
## Work hard

Start very early and do little else

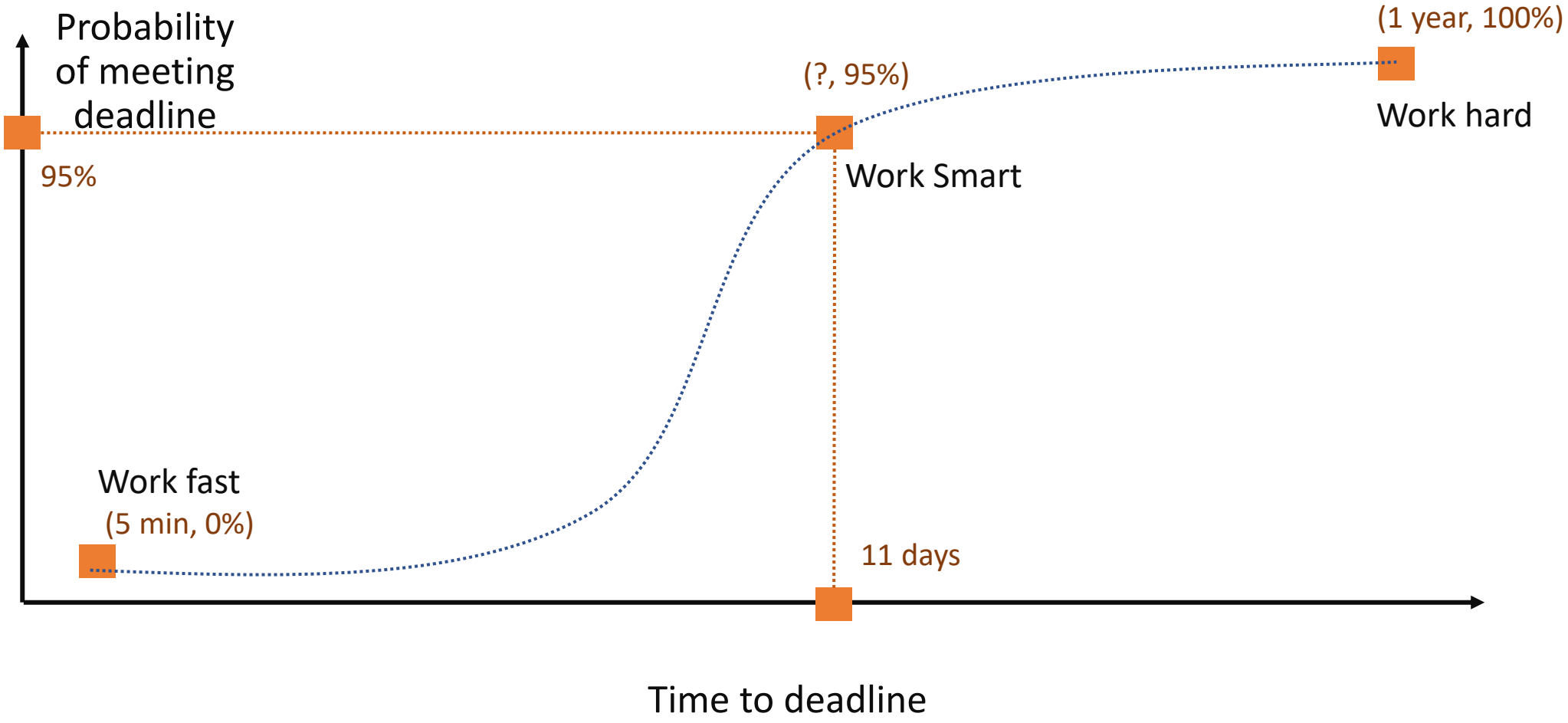
# Working Hard, Fast, Smart



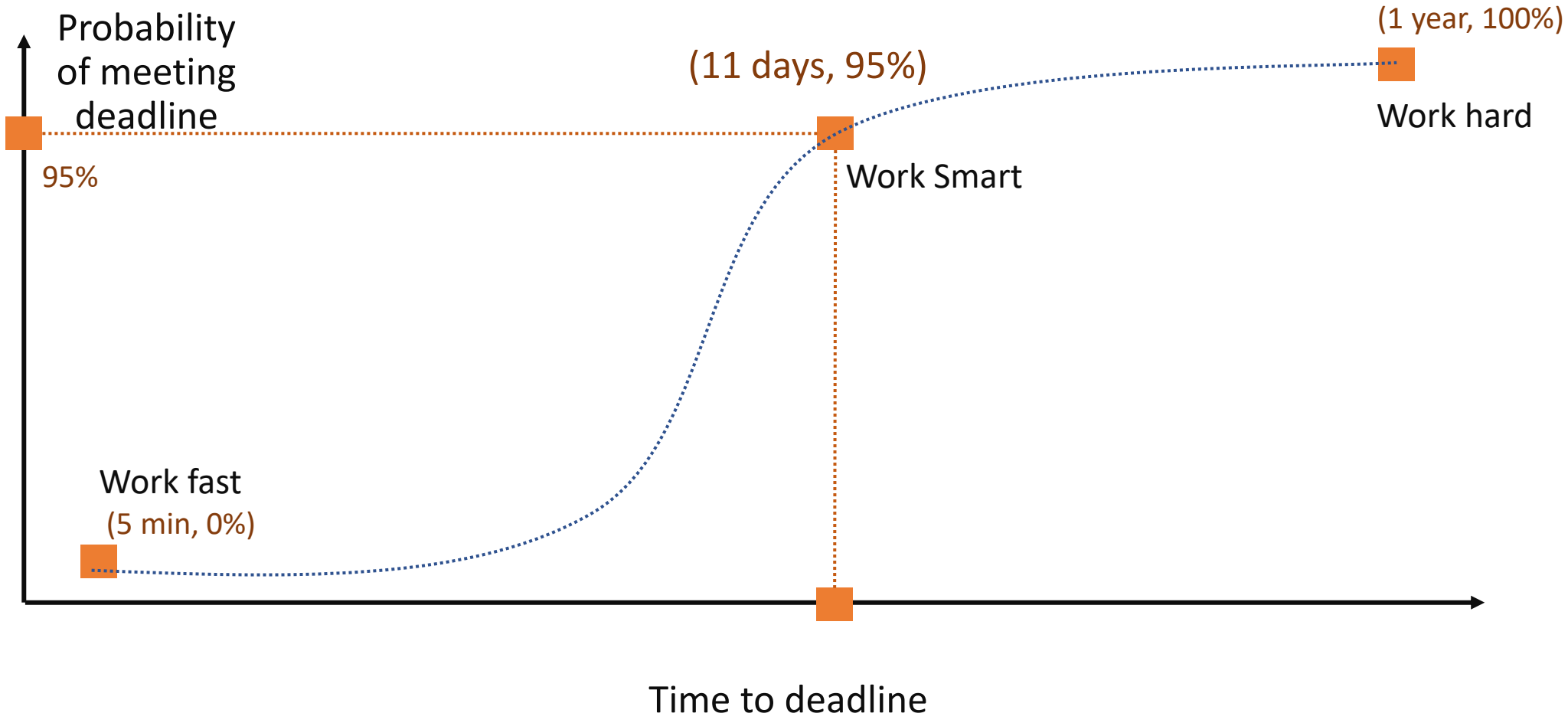
# Working Hard, Fast, Smart



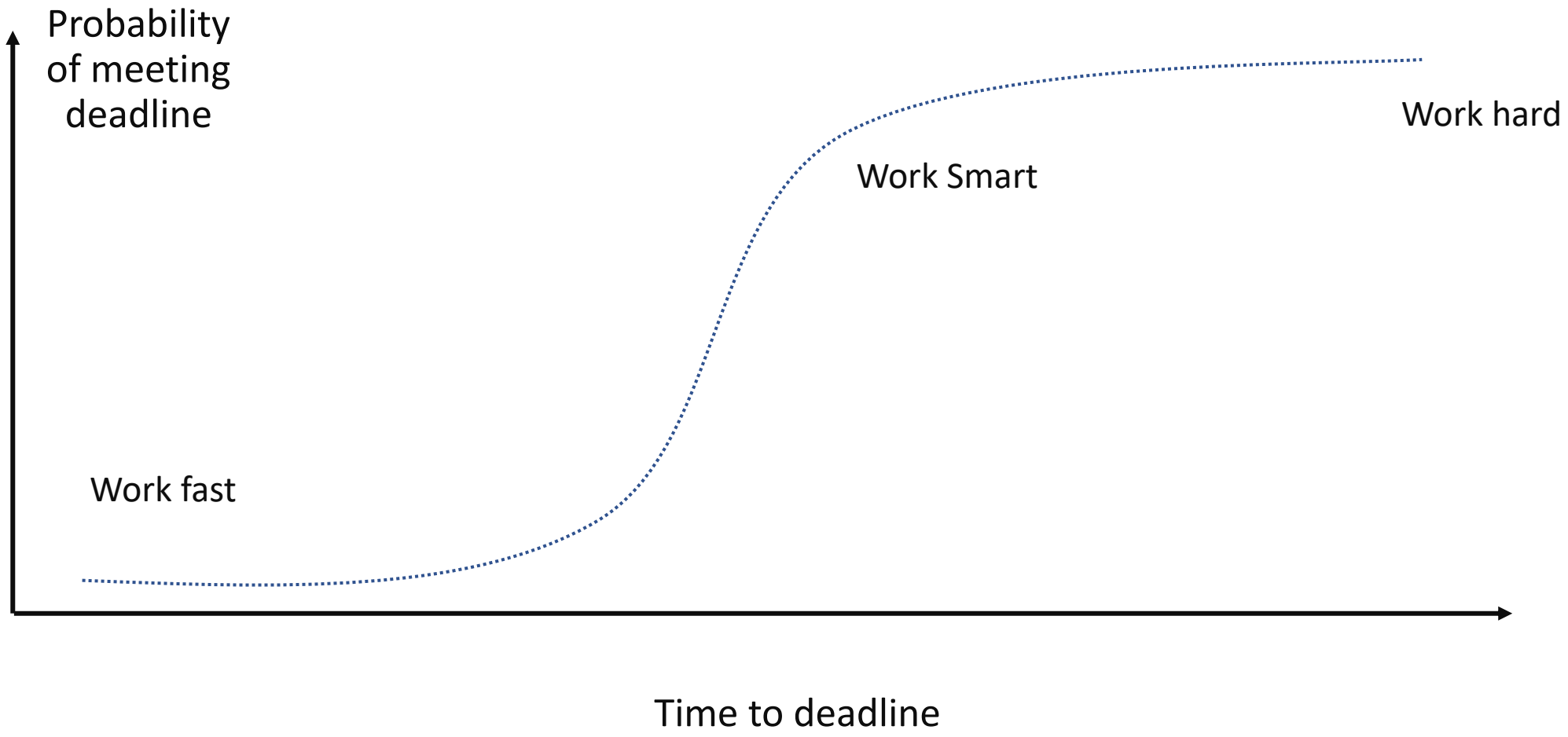
# Working Hard, Fast, Smart



# Working Hard, Fast, Smart

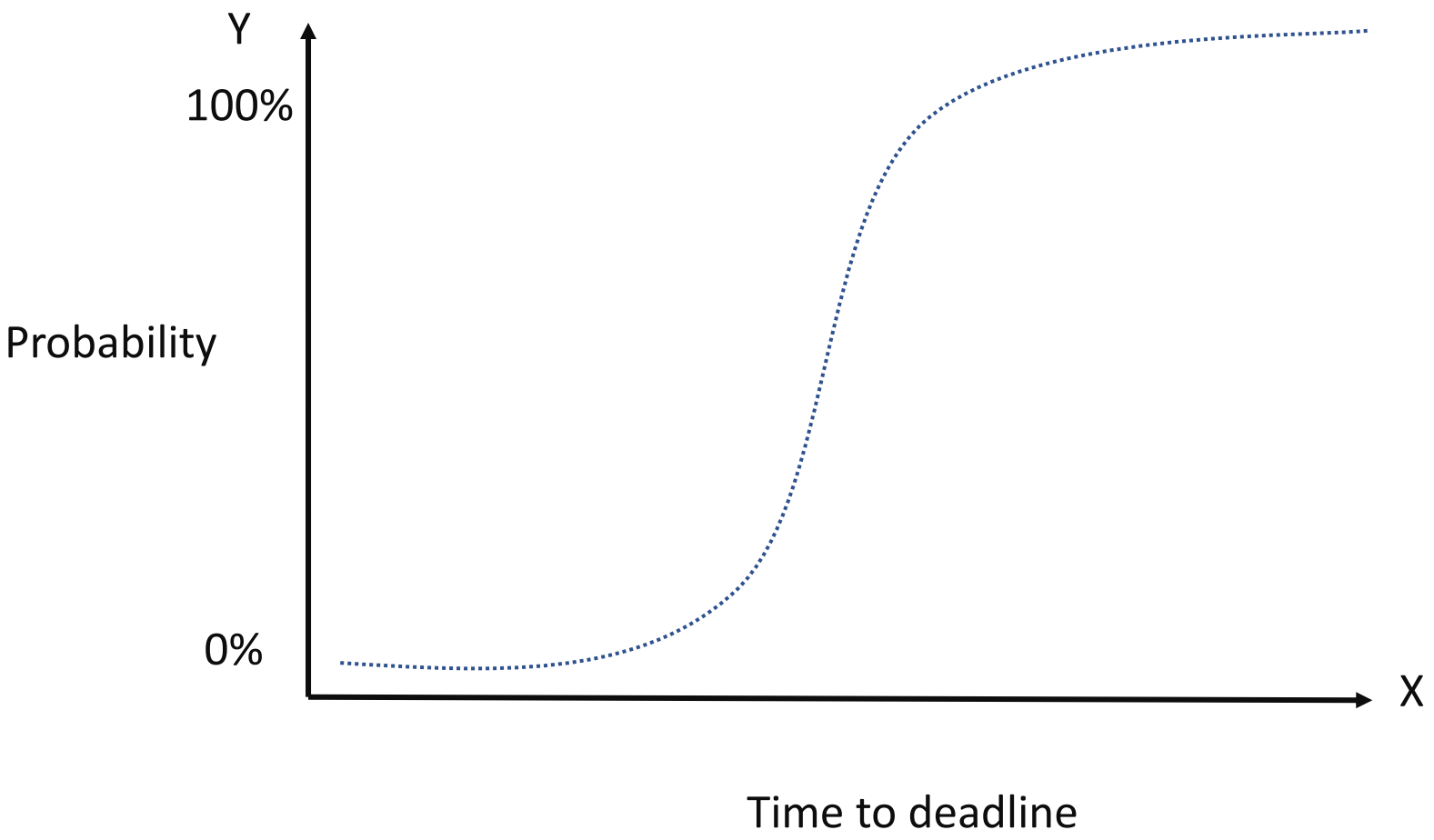


# Working Hard, Fast, Smart

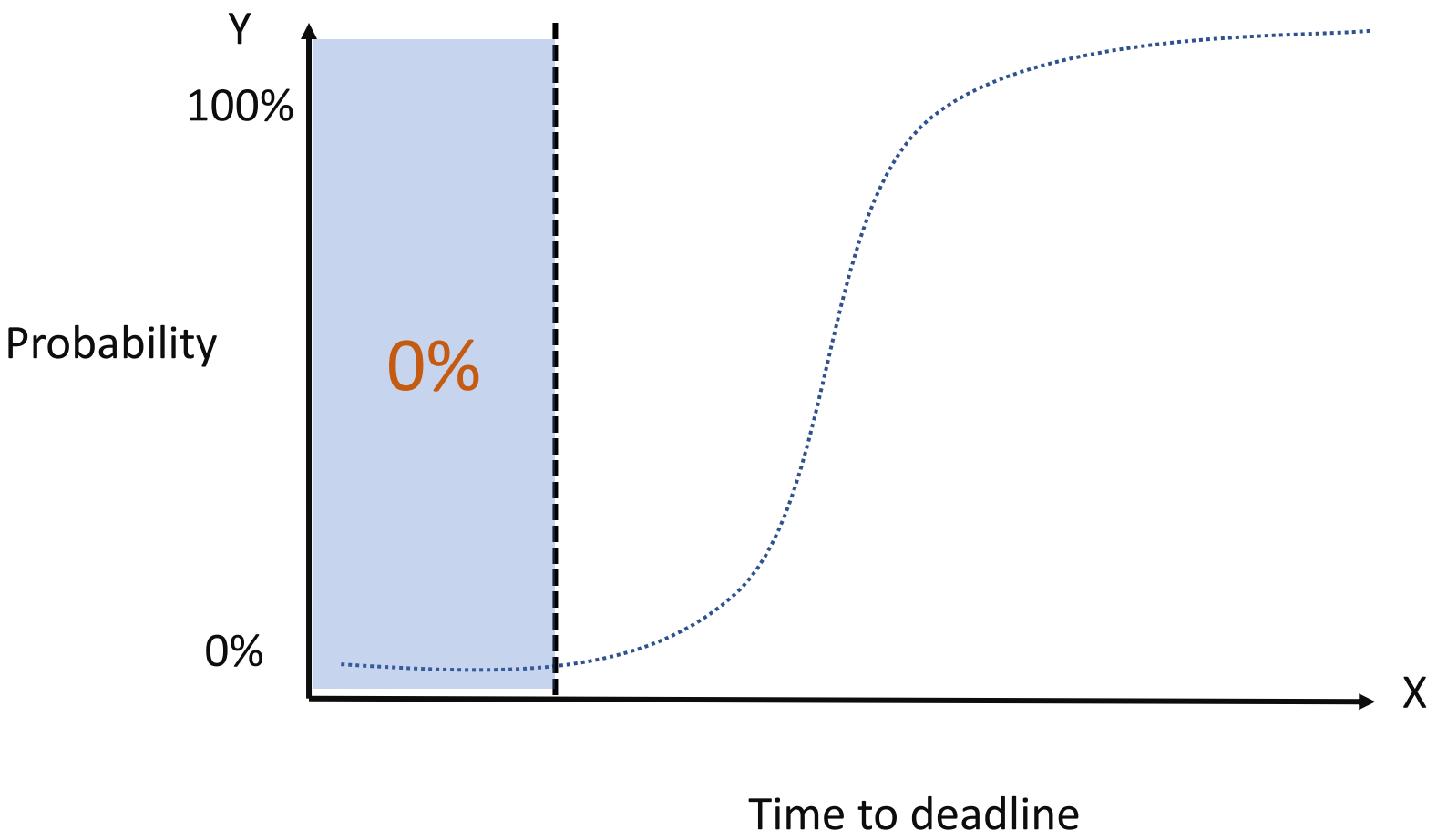


Logistic Regression helps find how probabilities are changed by action

# Working Smart with Logistic Regression

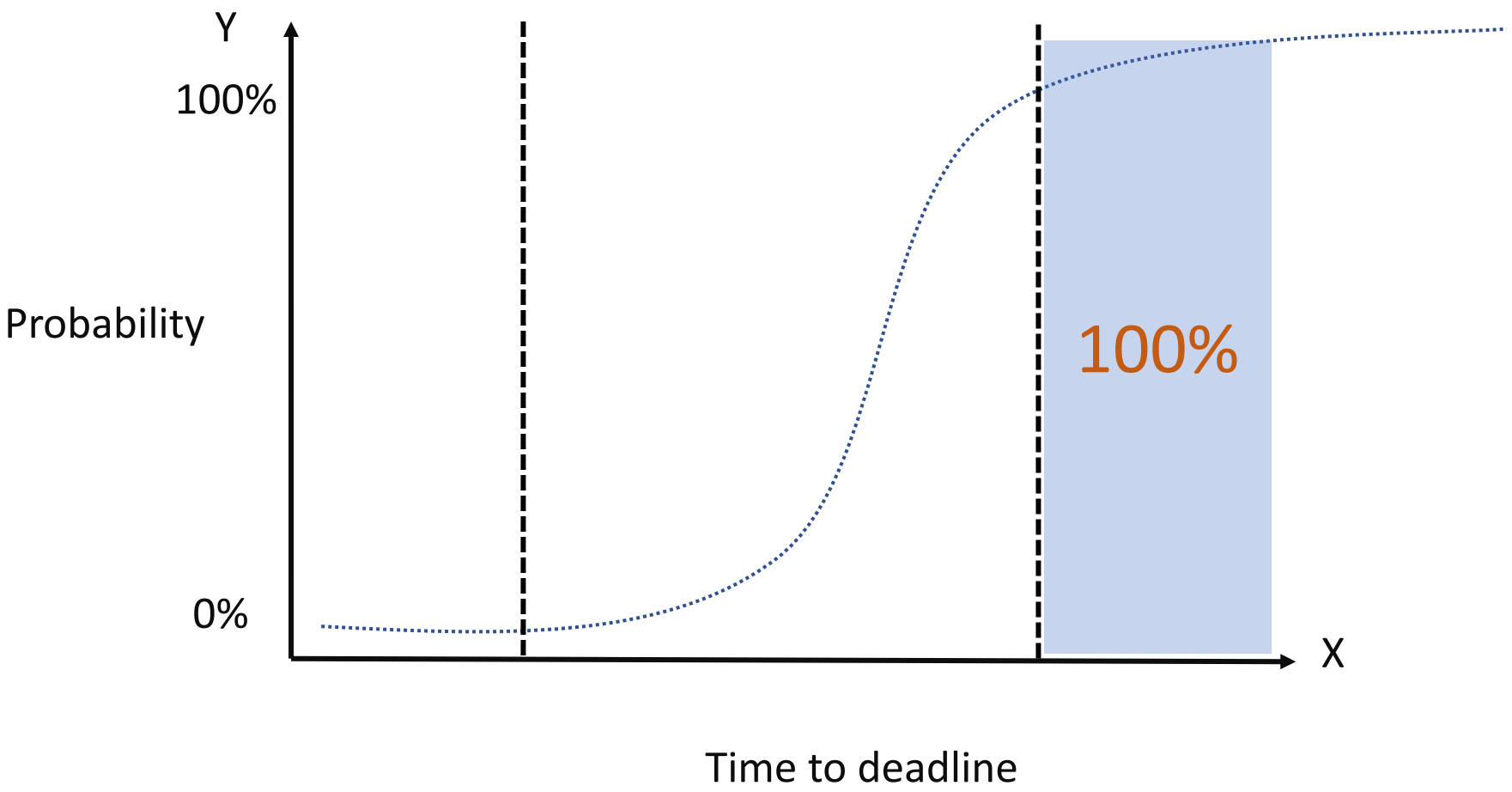


# Working Smart with Logistic Regression



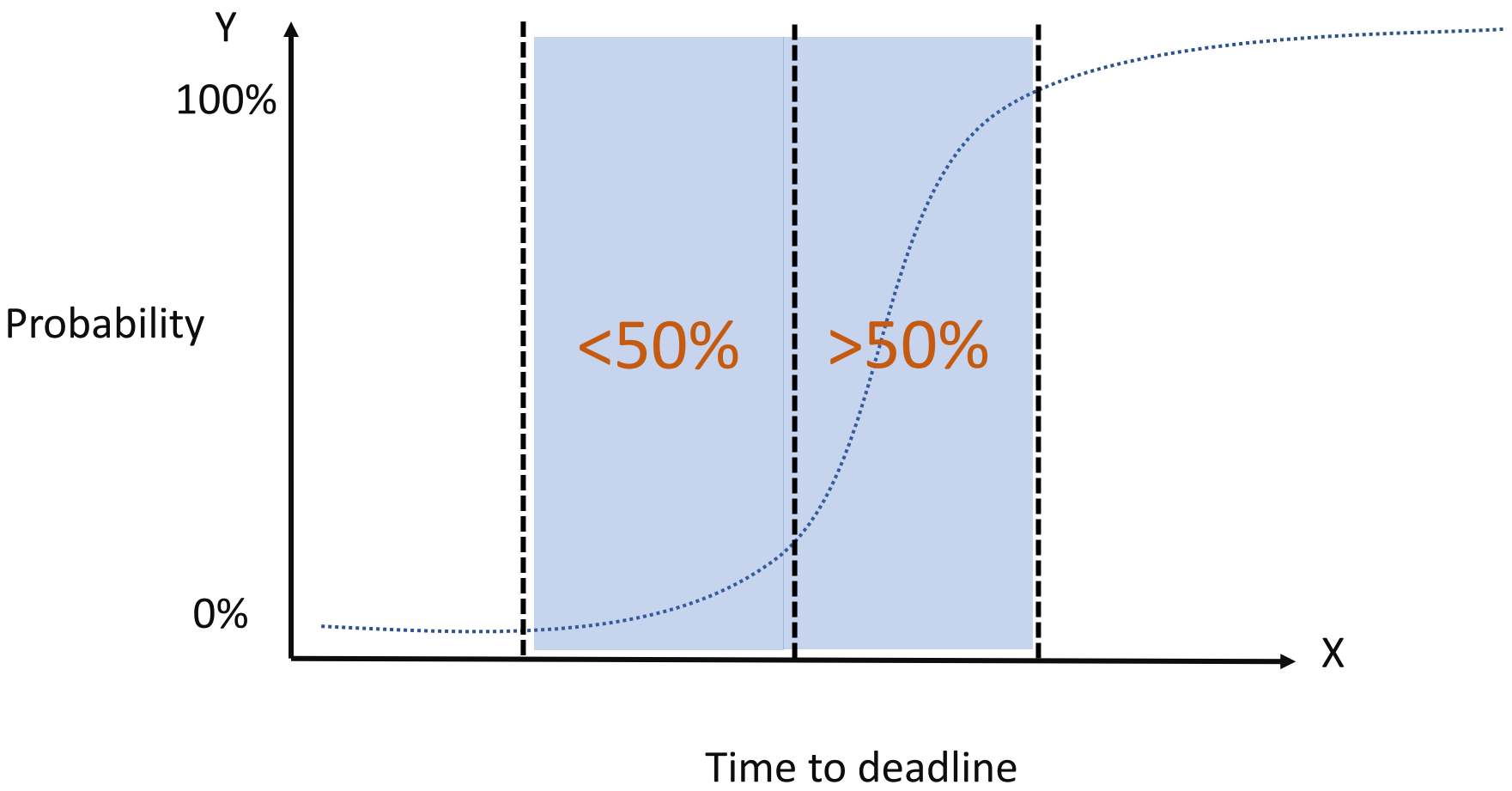

Start too late, and you'll definitely miss

# Working Smart with Logistic Regression

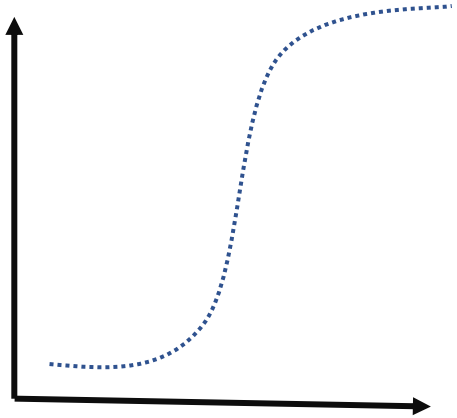



Start too early, and you'll definitely make it

# Working Smart with Logistic Regression




Working smart is knowing when to start



y: Hit or miss? (0 or 1?)

x: Start time before deadline

p(y) : Probability of y = 1

$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

**Logistic regression involves finding the "best fit" such curve**

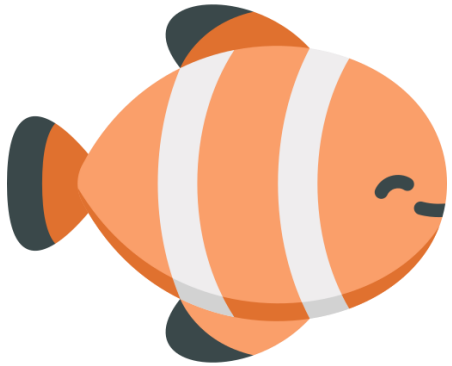
- A is the intercept
- B is the regression coefficient

**(e is the constant 2.71828)**

# Whales: Fish or Mammals?

## Mammals

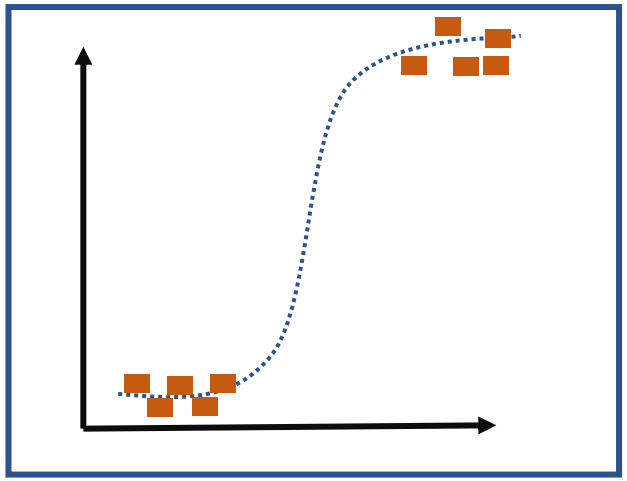
Members of the infraorder Cetacea



## Fish

Look like fish, swim like fish, and move like fish

# ML-based Predictor



Corpus

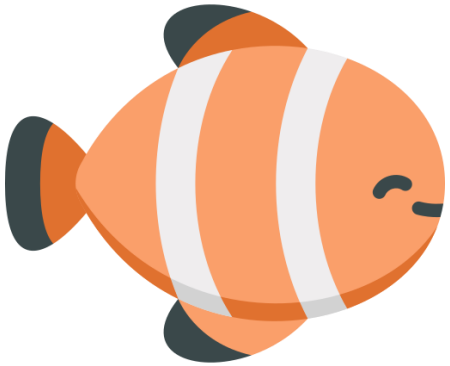
Logistic regression

ML-based predictor

$$p(y_i) = \frac{1}{1 + e^{-(A+Bx_i)}}$$

# Applying Logistic Regression

Mammals

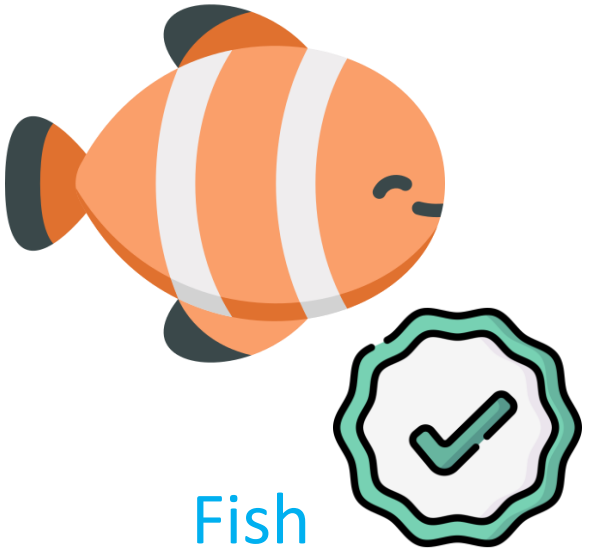


Fish

Probability of whales being fish < $P_{threshold}$

# Applying Logistic Regression

Mammals



Fish

Probability of whales being fish > $P_{threshold}$