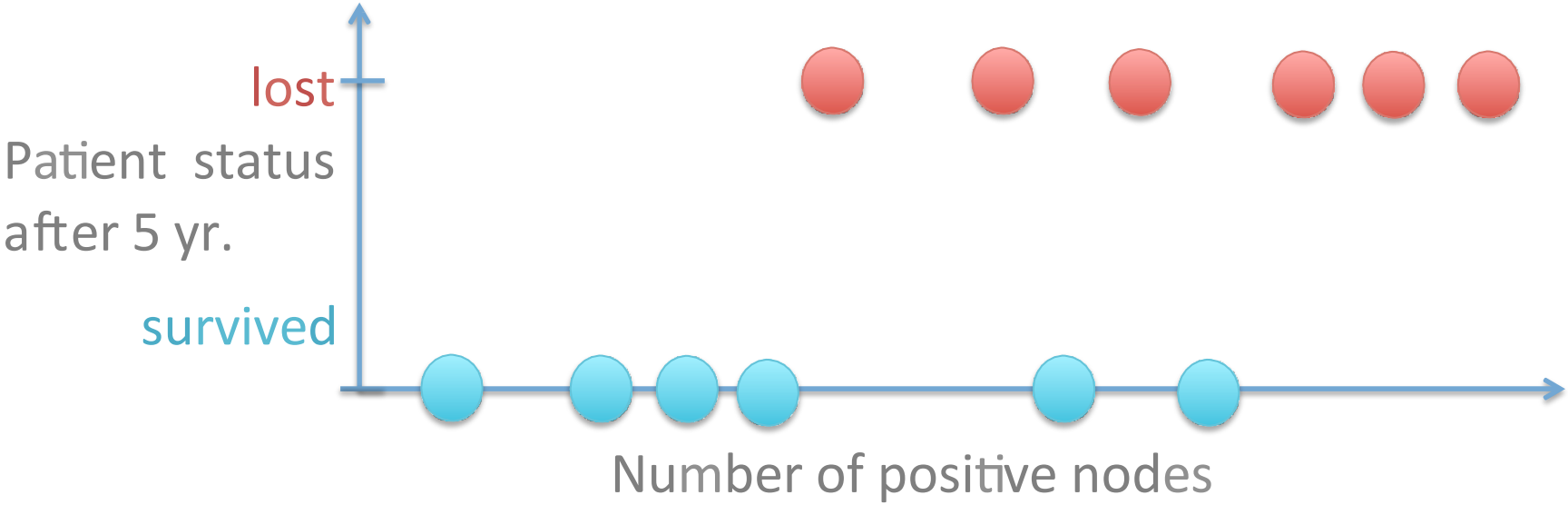
Logistic Regression

1 Feature: Number of + nodes
2 Labels: Survived / Lost



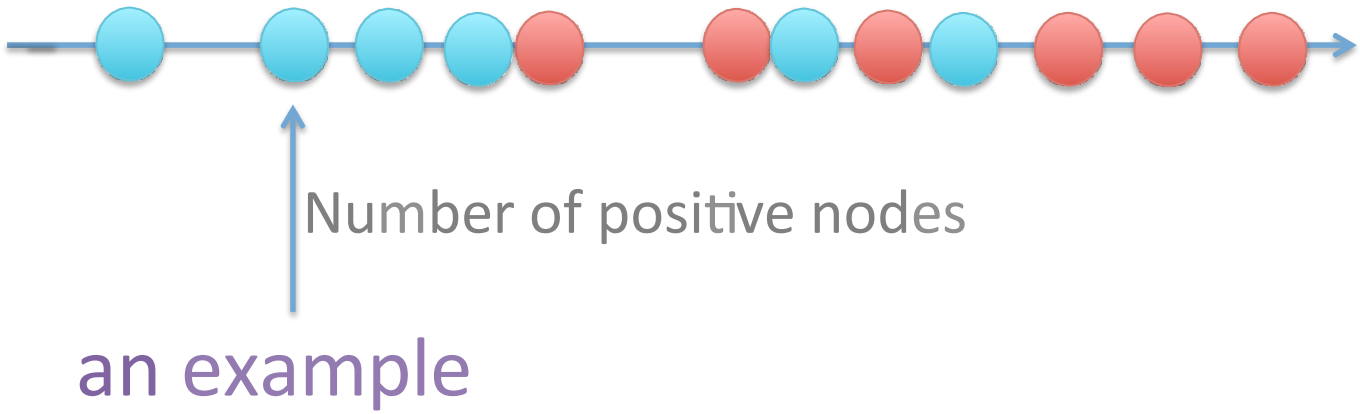
1 Feature: Number of + nodes

2 Labels: Survived / Lost

Number of positive nodes

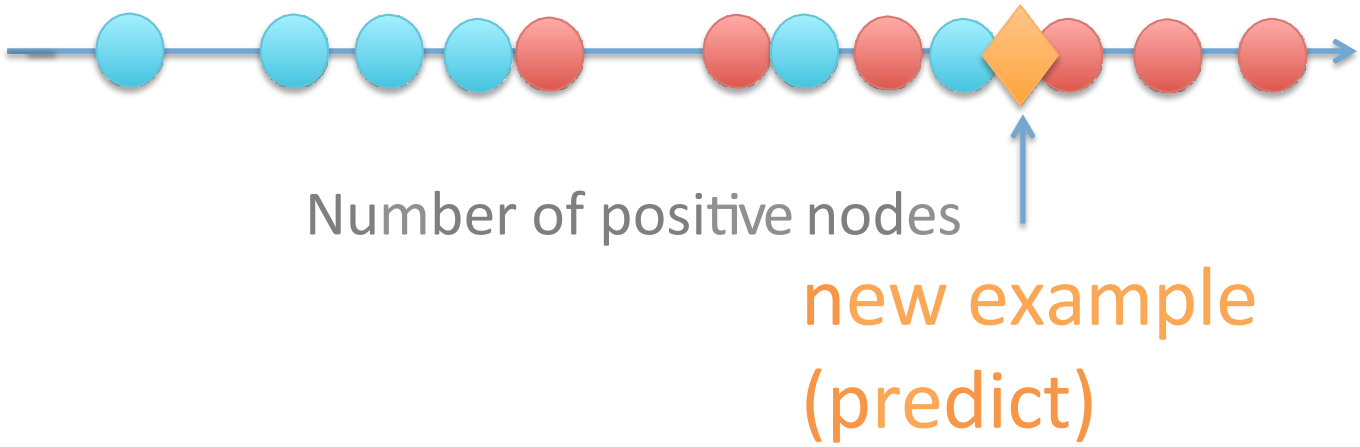
1 Feature: Number of + nodes

2 Labels: Survived / Lost



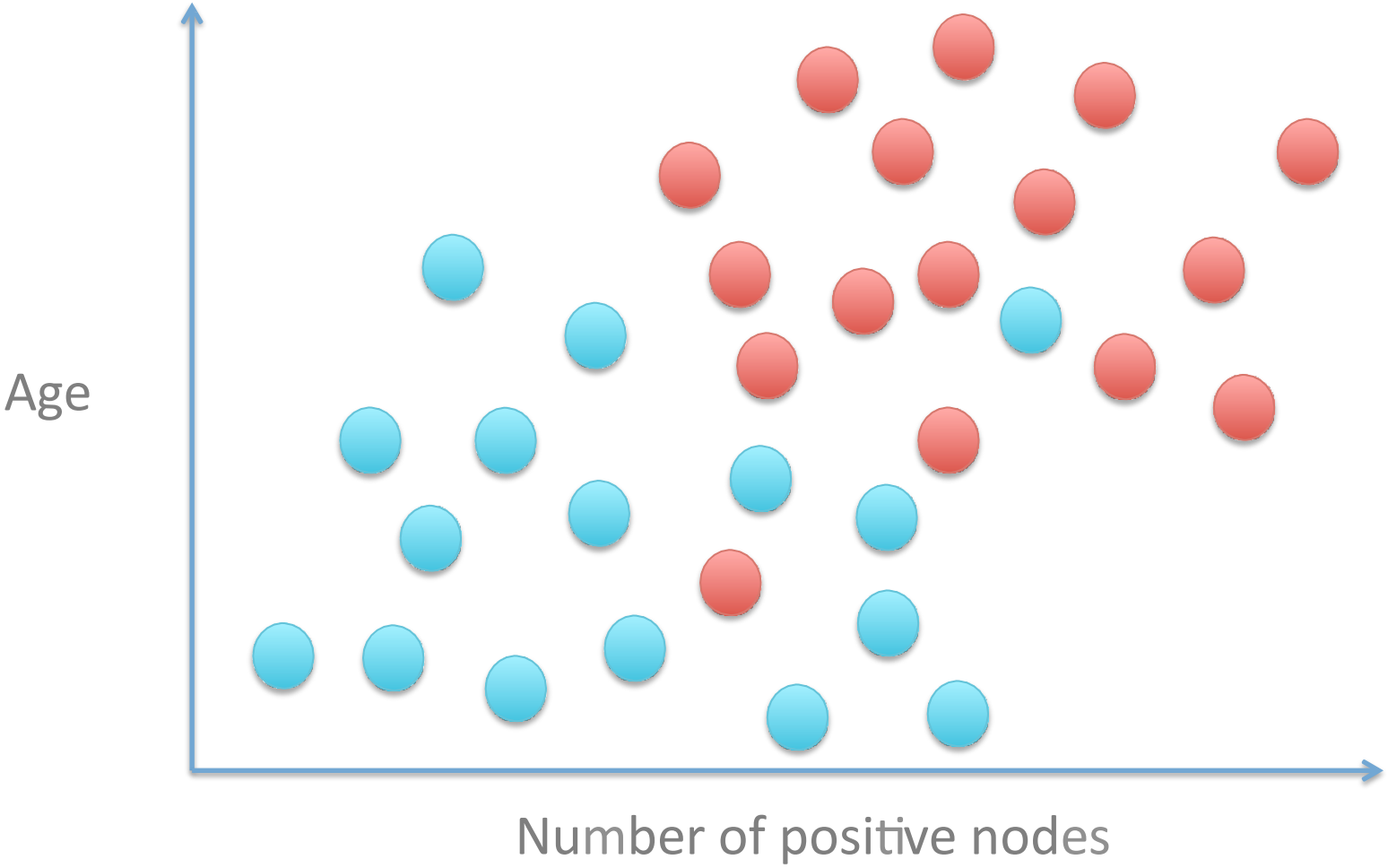
1 Feature: Number of + nodes

2 Labels: Survived / Lost



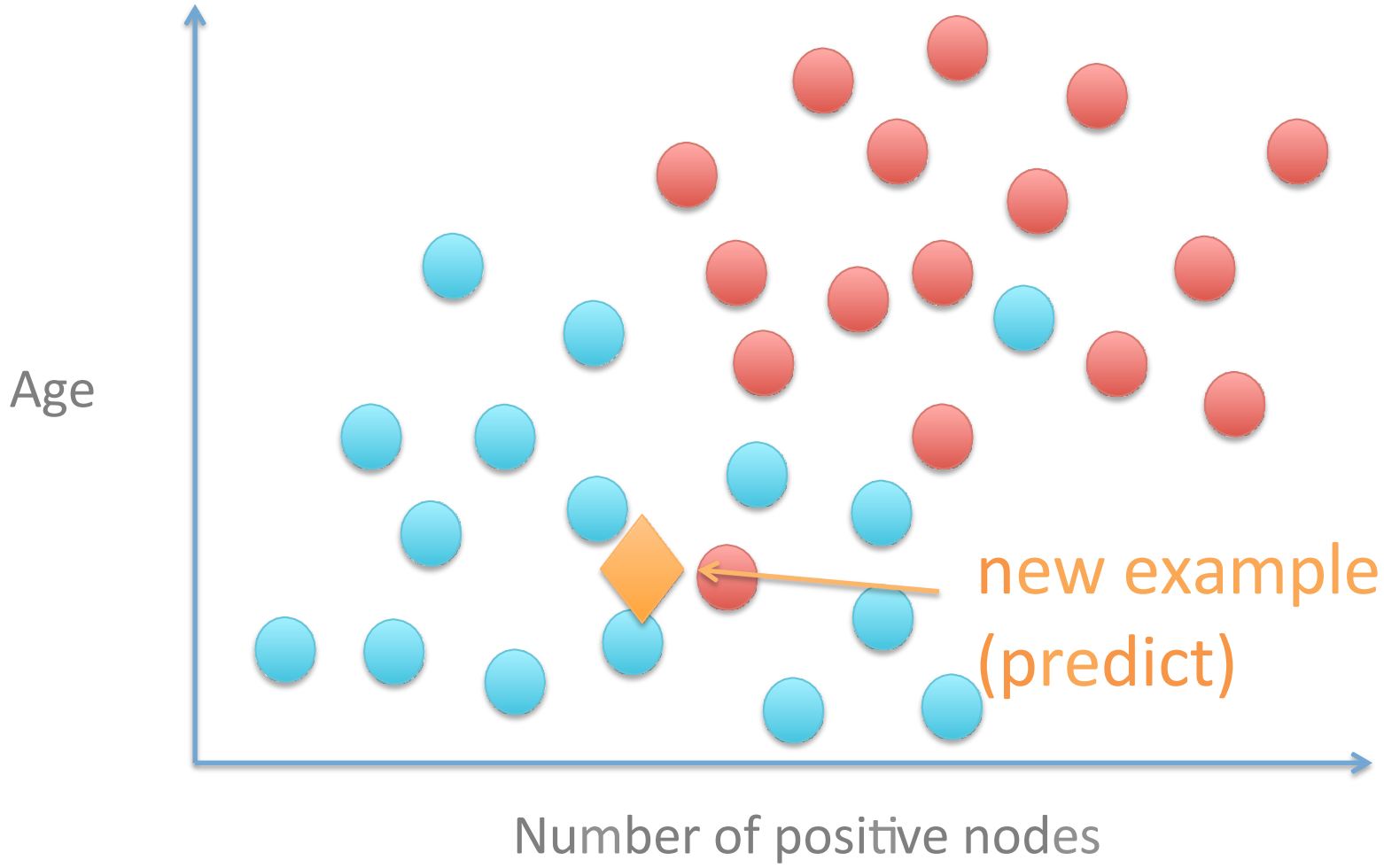
2 Features: Number of + nodes, Age

2 Labels: Survived / Lost



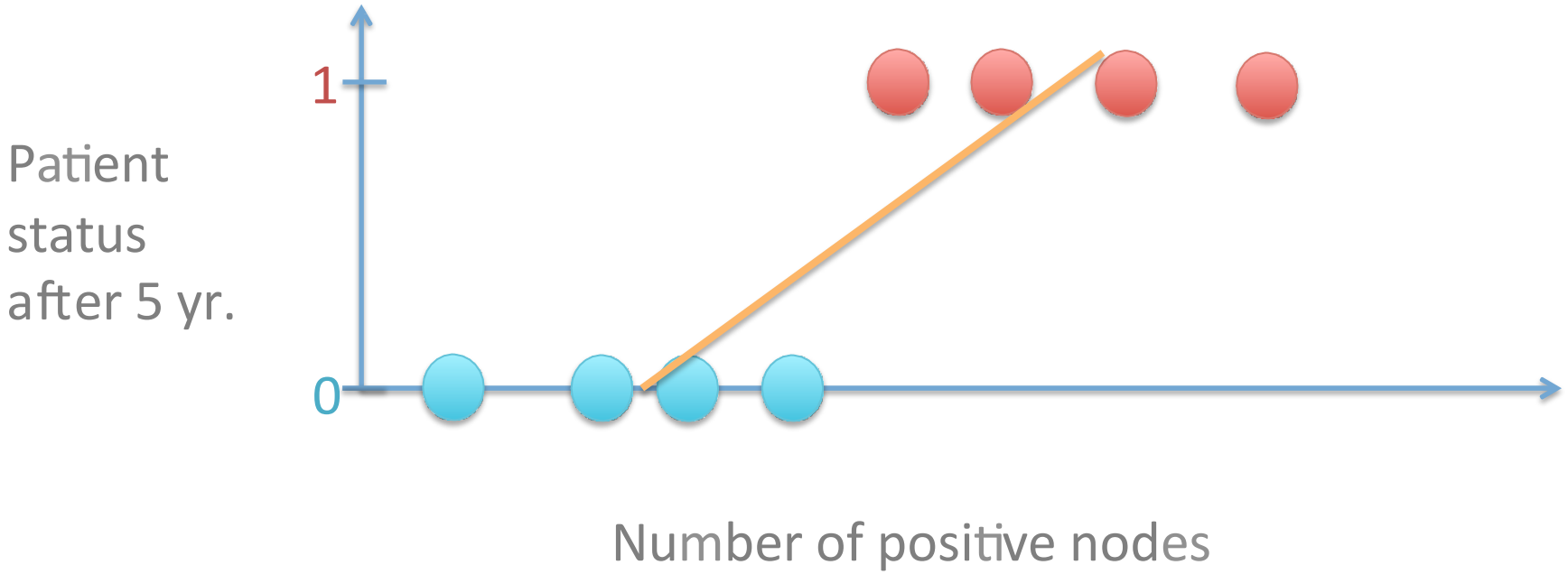
2 Features: Number of + nodes, Age

2 Labels: Survived / Lost



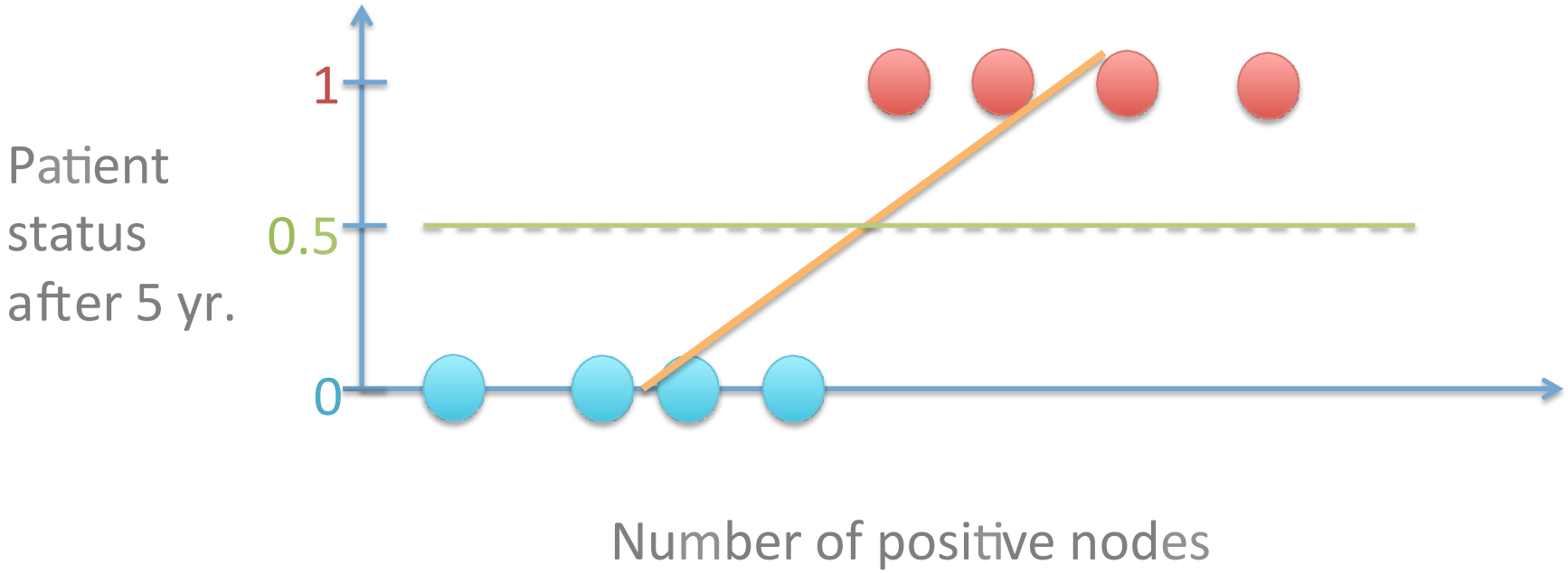
Linear regression for classification?

Linear regression for classification?



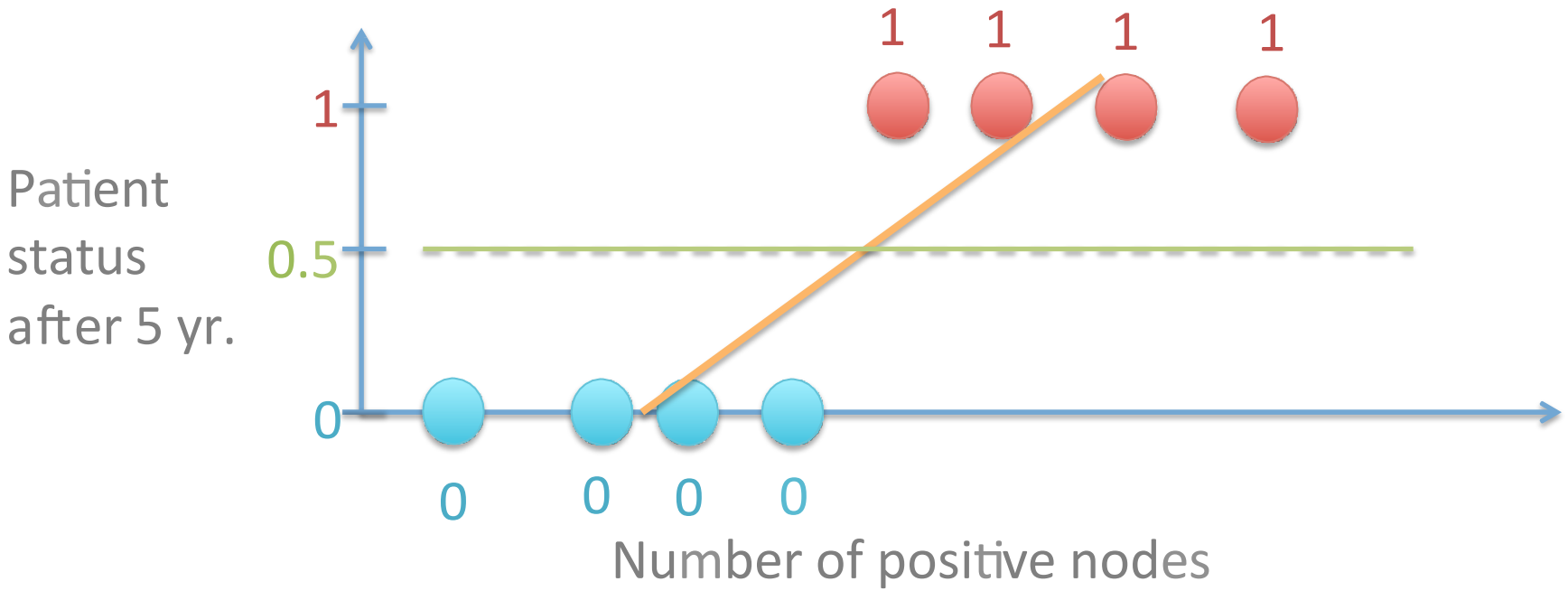
$$y_\beta(x) = \beta_0 + \beta_1 x + \varepsilon$$

Linear regression for classification?



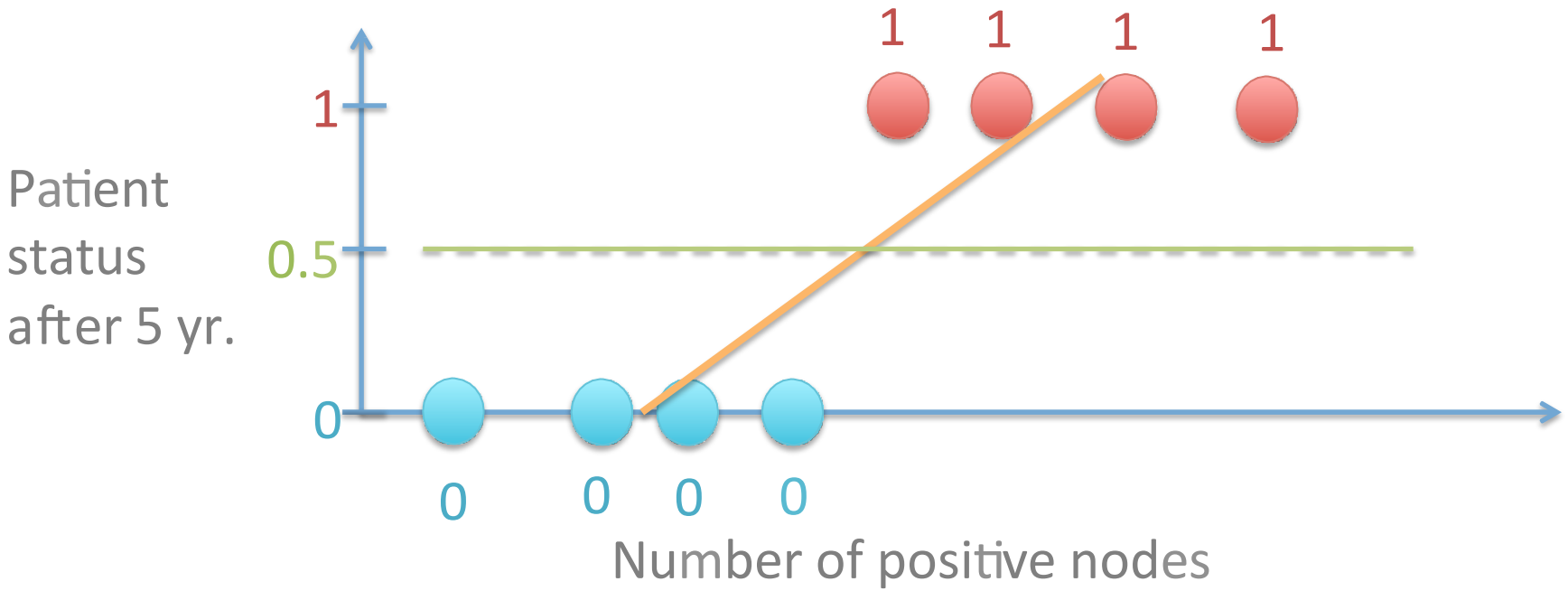
If y_pred > 0.5 predict label 1 (lost)
If y_pred < 0.5 predict label 0 (survived)

Linear regression for classification?



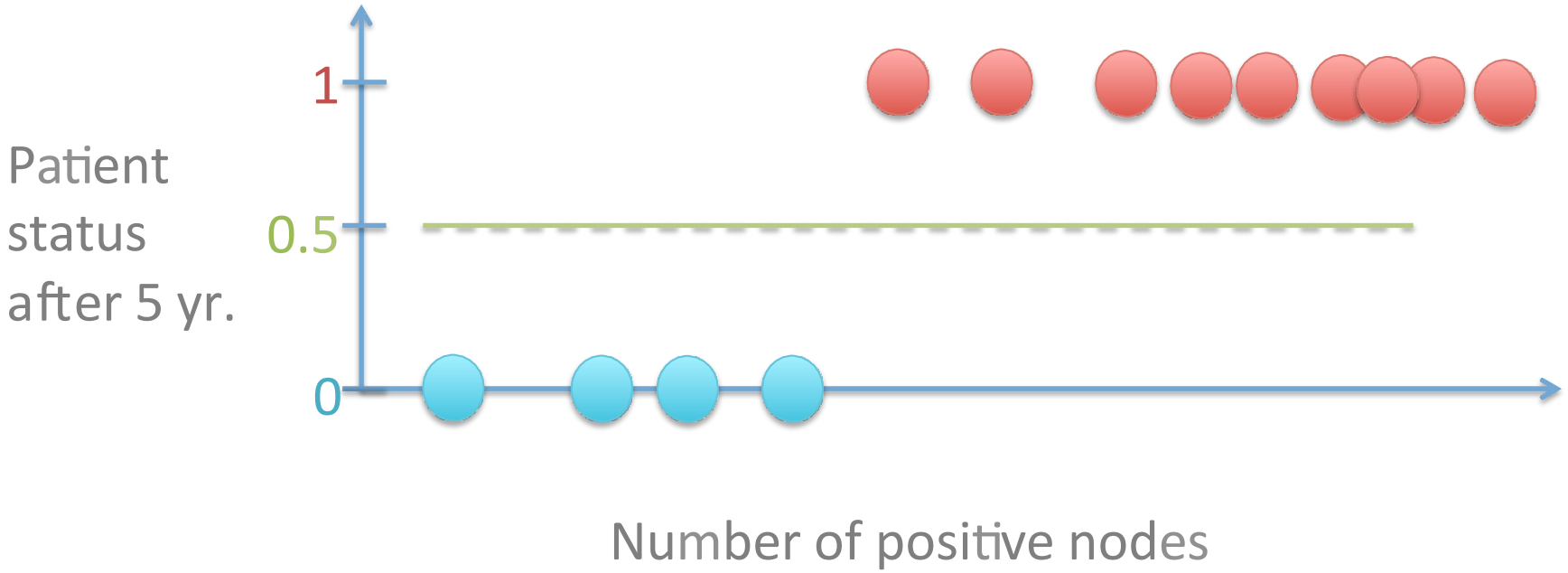
If y_pred > 0.5 predict label 1 (lost)
If y_pred < 0.5 predict label 0 (survived)

Linear regression for classification?



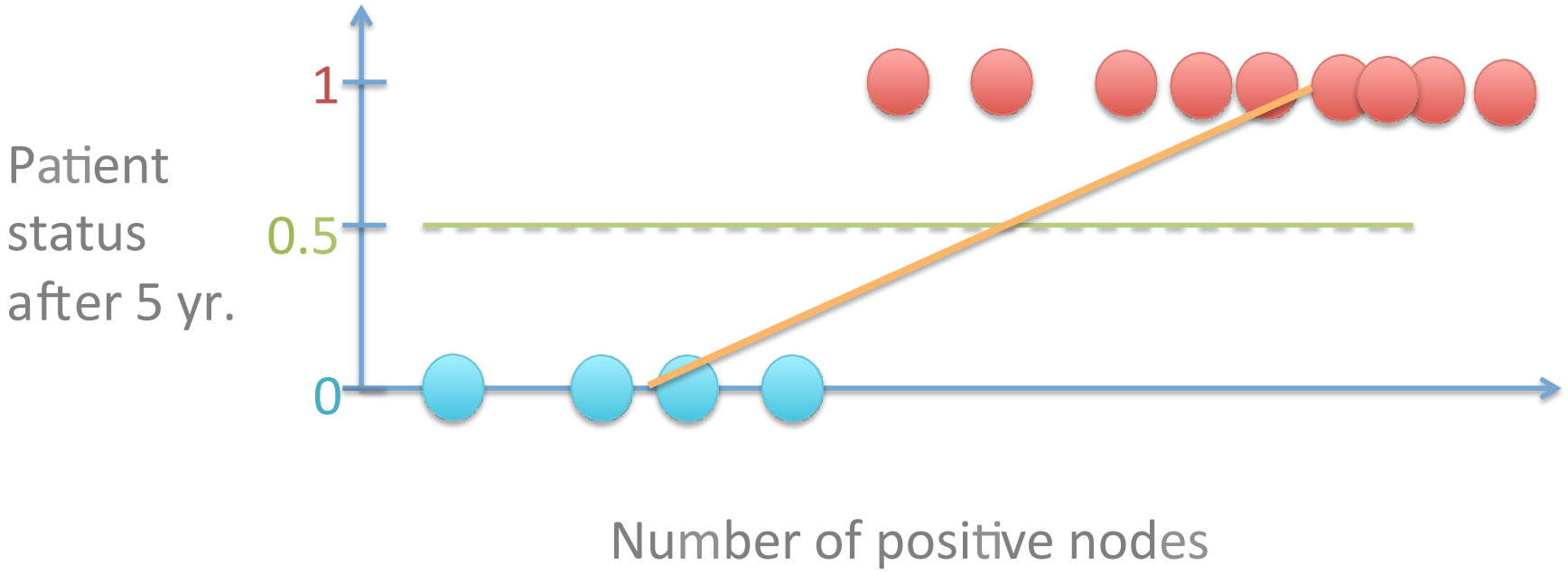
If y_pred > 0.5 predict label 1 (lost)
If y_pred < 0.5 predict label 0 (survived)

Linear regression for classification?



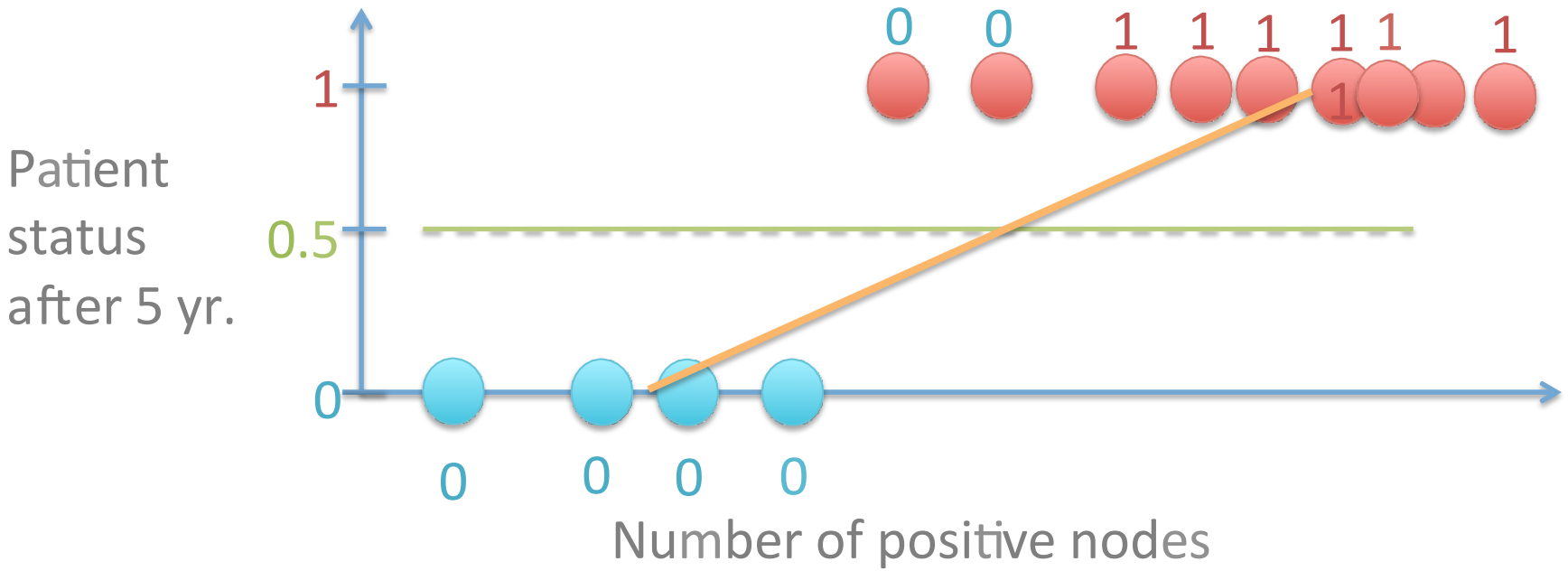
If y_pred > 0.5 predict label 1 (lost)
If y_pred < 0.5 predict label 0 (survived)

Linear regression for classification?



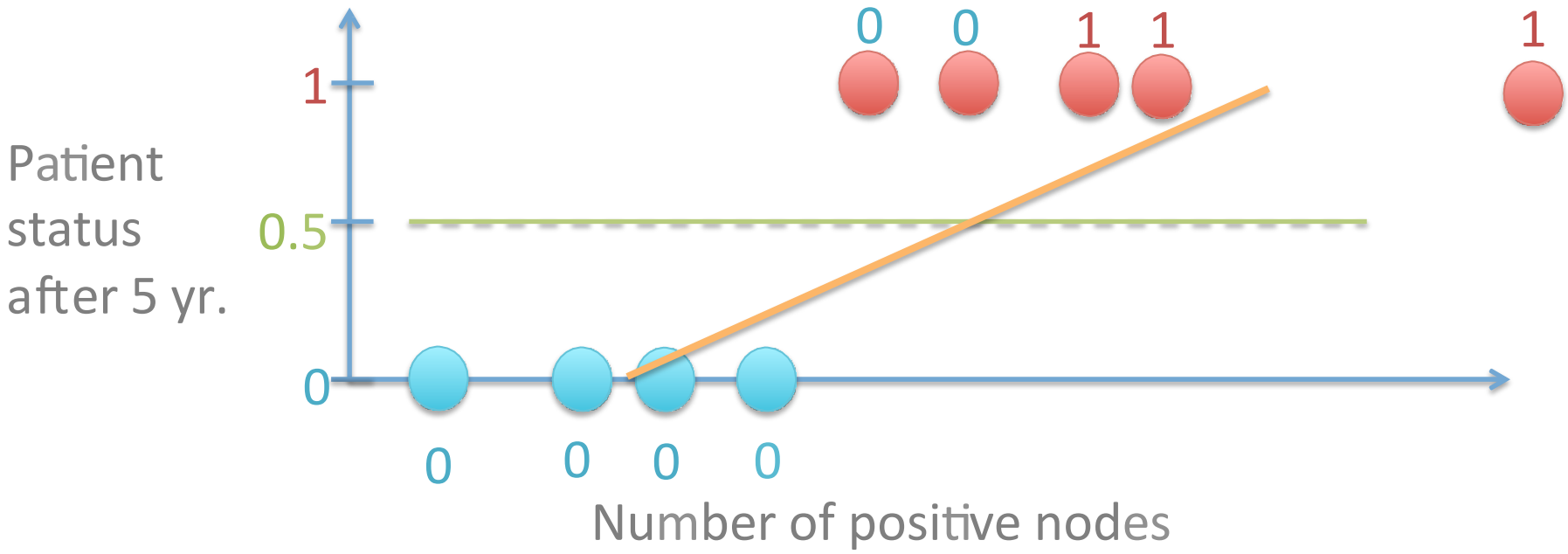
If y_pred > 0.5 predict label 1 (lost)
If y_pred < 0.5 predict label 0 (survived)

Linear regression for classification?



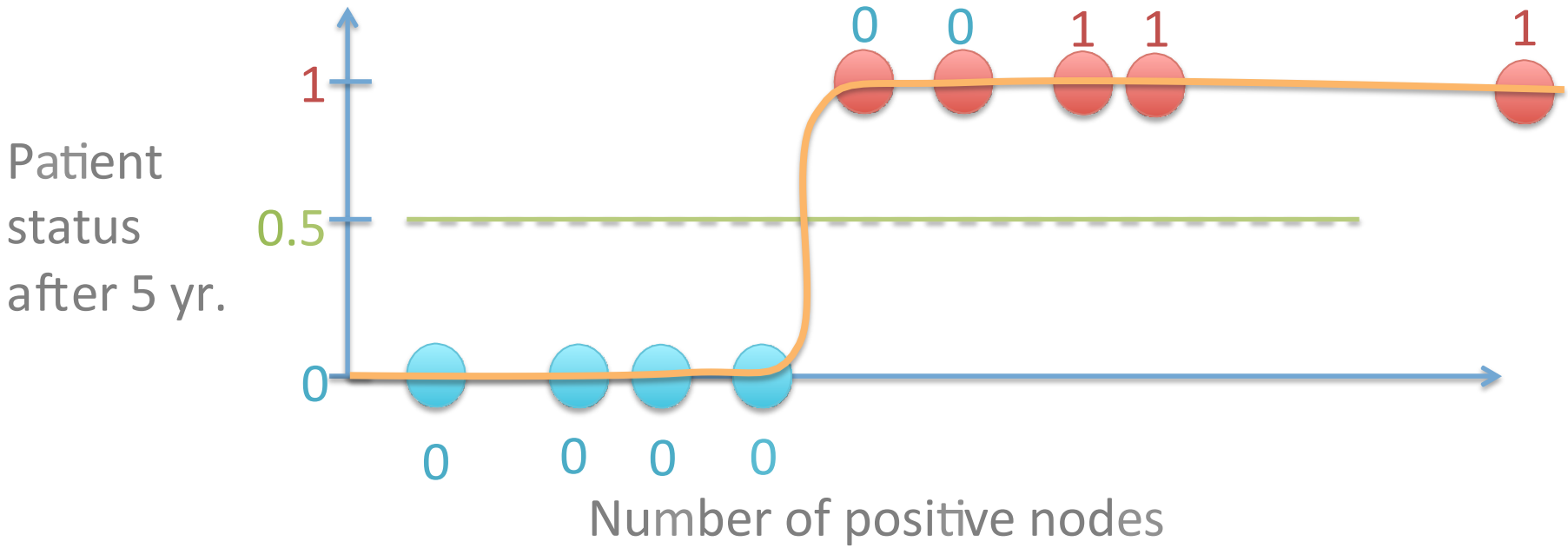
If y_pred > 0.5 predict label 1 (lost)
If y_pred < 0.5 predict label 0 (survived)

Linear regression for classification?



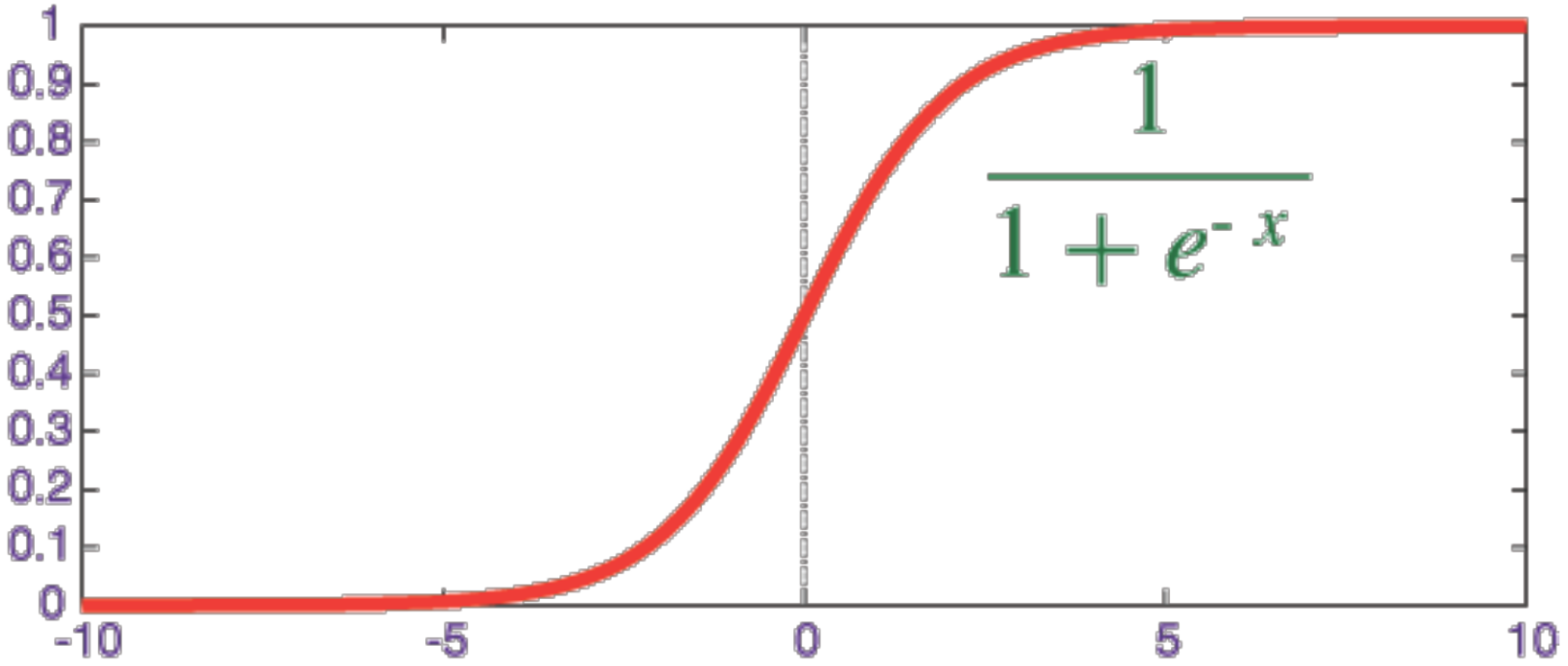
If y_pred > 0.5 predict label 1 (lost)
If y_pred < 0.5 predict label 0 (survived)

Logistic regression to the rescue

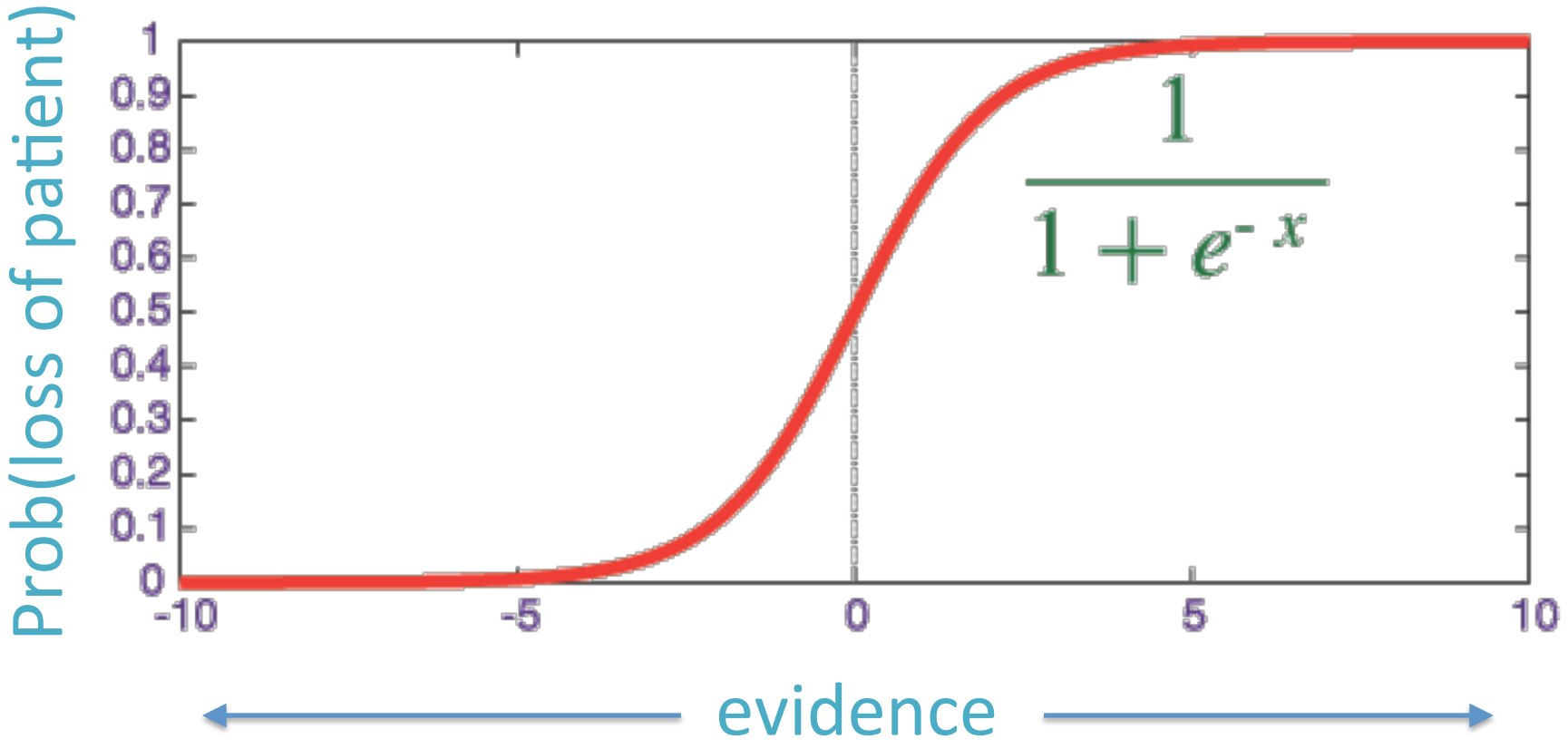


$$y_\beta(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x+\varepsilon)}}$$

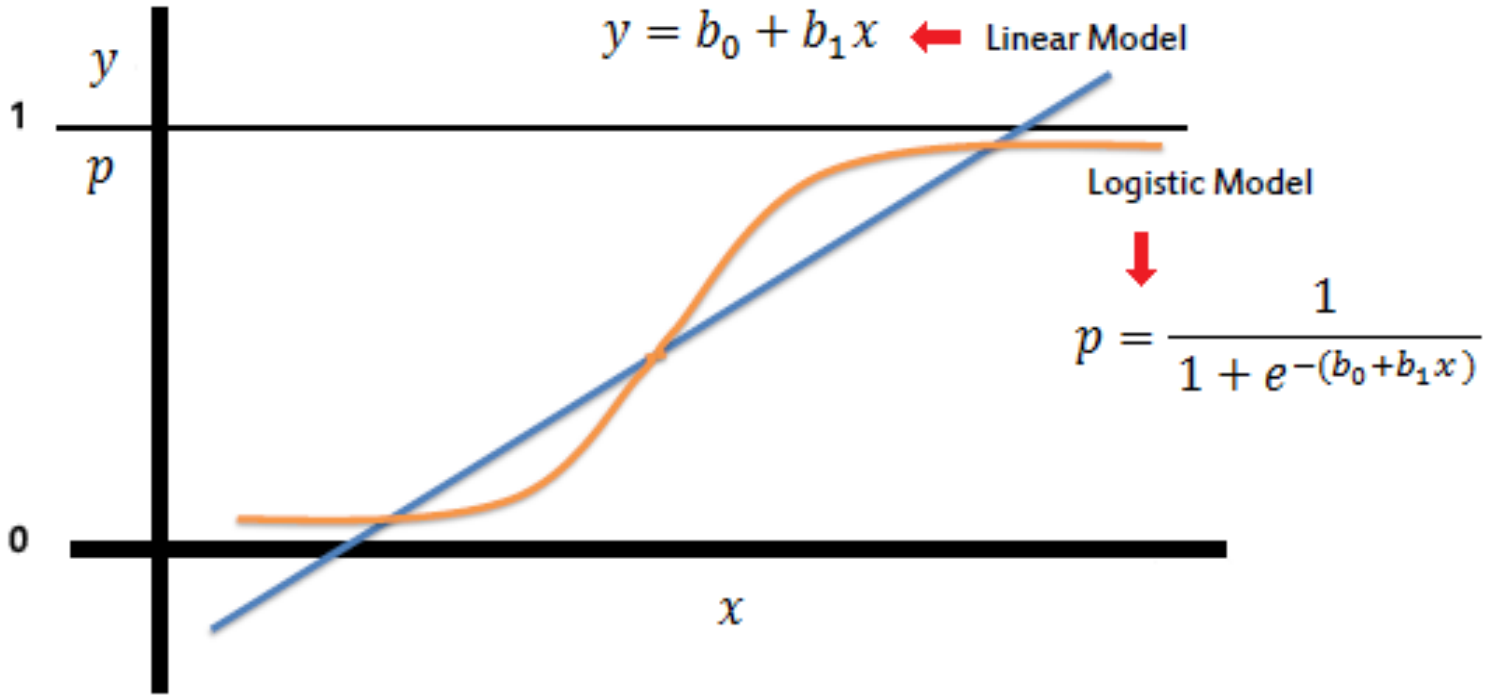
What is this function?



What is this function?

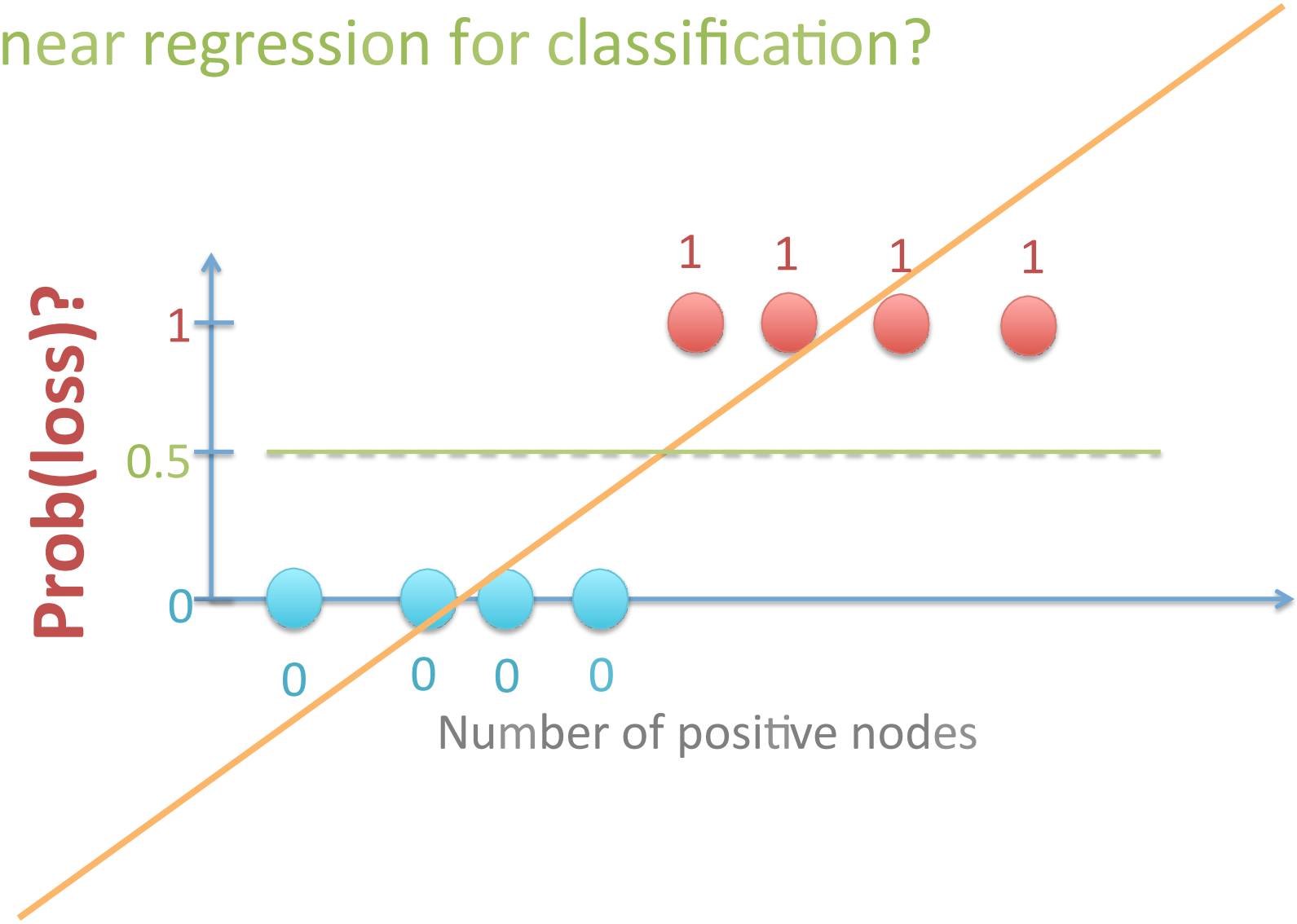


What is this function?

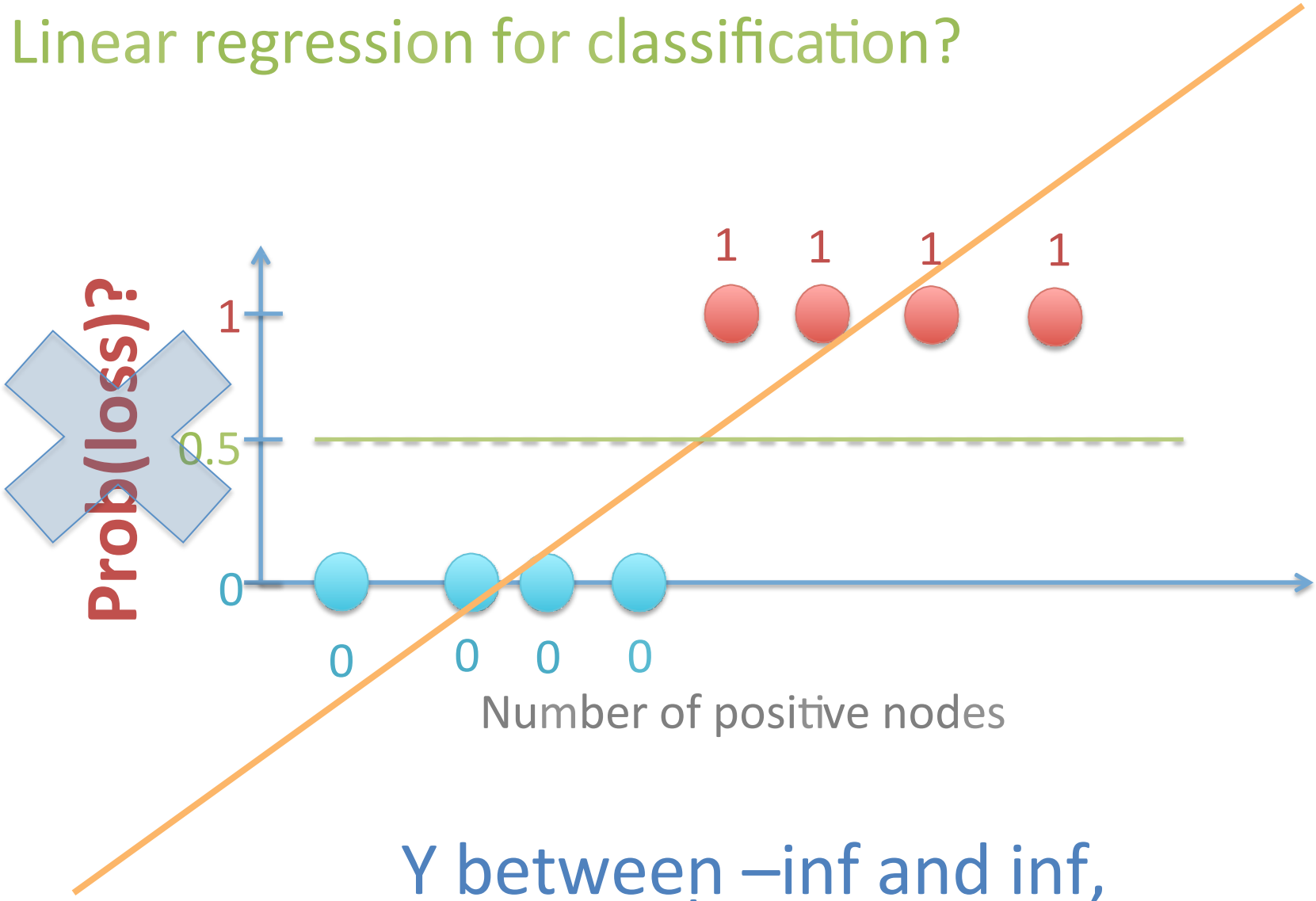


http://www.saedsayad.com/logistic_regression.htm

Linear regression for classification?



Linear regression for classification?



Y between –inf and inf,
Not 0 and 1

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.8$$

$$P(survival) = 0.2$$

Probability

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.8$$

$$P(survival) = 0.2$$

Probability

$$\frac{P(loss)}{P(survival)} = 4$$

Odds

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.05$$
$$P(survival) = 0.95$$

Probability

$$\frac{P(loss)}{P(survival)} = 0.053$$

Odds

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.5$$

$$P(survival) = 0.5$$

Probability

$$\frac{P(loss)}{P(survival)} = 1$$

Odds

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.5$$

$$P(survival) = 0.5$$

Probability

$$\frac{P(loss)}{P(survival)} = 1$$

Odds

between 0 and inf

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.5$$

$$P(survival) = 0.5$$

Probability

$$\log\left(\frac{P(loss)}{P(survival)}\right) = 0$$

Log Odds

between -inf and inf

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.05$$
$$P(survival) = 0.95$$

Probability

$$\log\left(\frac{P(loss)}{P(survival)}\right) = -2.94$$

Log Odds

between -inf and inf

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.8$$
$$P(survival) = 0.2$$

Probability

$$\log\left(\frac{P(loss)}{P(survival)}\right) = 1.39$$

Log Odds

between -inf and inf

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.999$$
$$P(survival) = 0.001$$

Probability

$$\log\left(\frac{P(loss)}{P(survival)}\right) = 6.9$$

Log Odds

between -inf and inf

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.999$$

$$1 - P(loss) = 0.001$$

Probability

$$\log\left(\frac{P(loss)}{1 - P(loss)}\right) = 6.9$$

Log Odds

logit function

What metric would express the chances of loss/survival, but not constrained to [0, 1] ?

$$P(loss) = 0.999$$
$$1 - P(loss) = 0.001$$

Probability

$$\log\left(\frac{P(loss)}{1 - P(loss)}\right) = 6.9$$

Log Odds

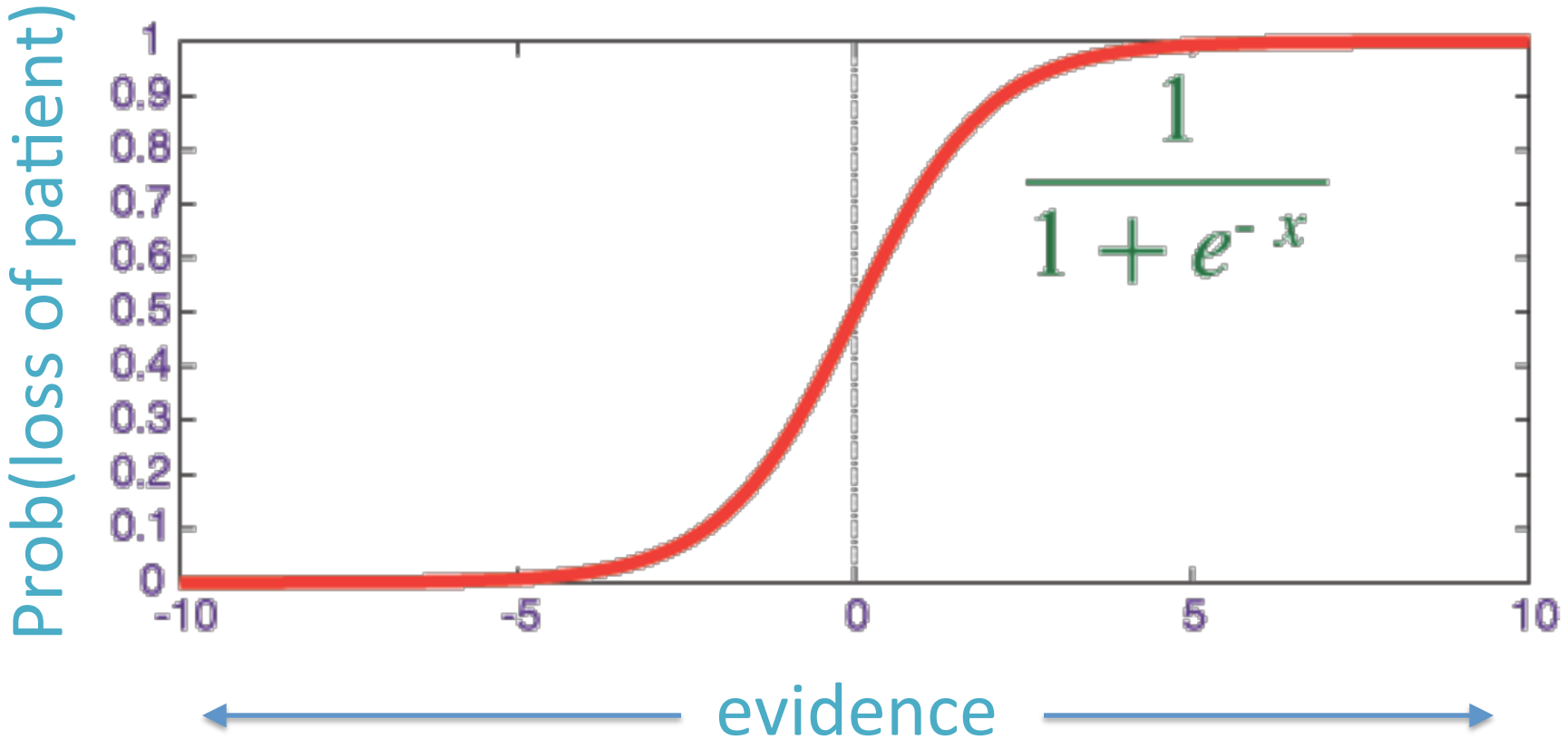
logit function

$$\frac{1}{1 + e^{-\log\left(\frac{P(loss)}{1-P(loss)}\right)}} = P(loss)$$

Logistic Function

Log Odds → Prob

What is this function?



"Logistic regression" is a classification algorithm

$$y_\beta(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x + \varepsilon)}}$$

```
from sklearn.linear_model import LogisticRegression
#(just like LinearRegression)
```

```
from statsmodels.formula.api import Logit
#(just like OLS)
```