Guided Tour of Machine Learning in Finance

Week 2-Lesson 3-part 2:

Probabilistic classification models

Igor Halperin

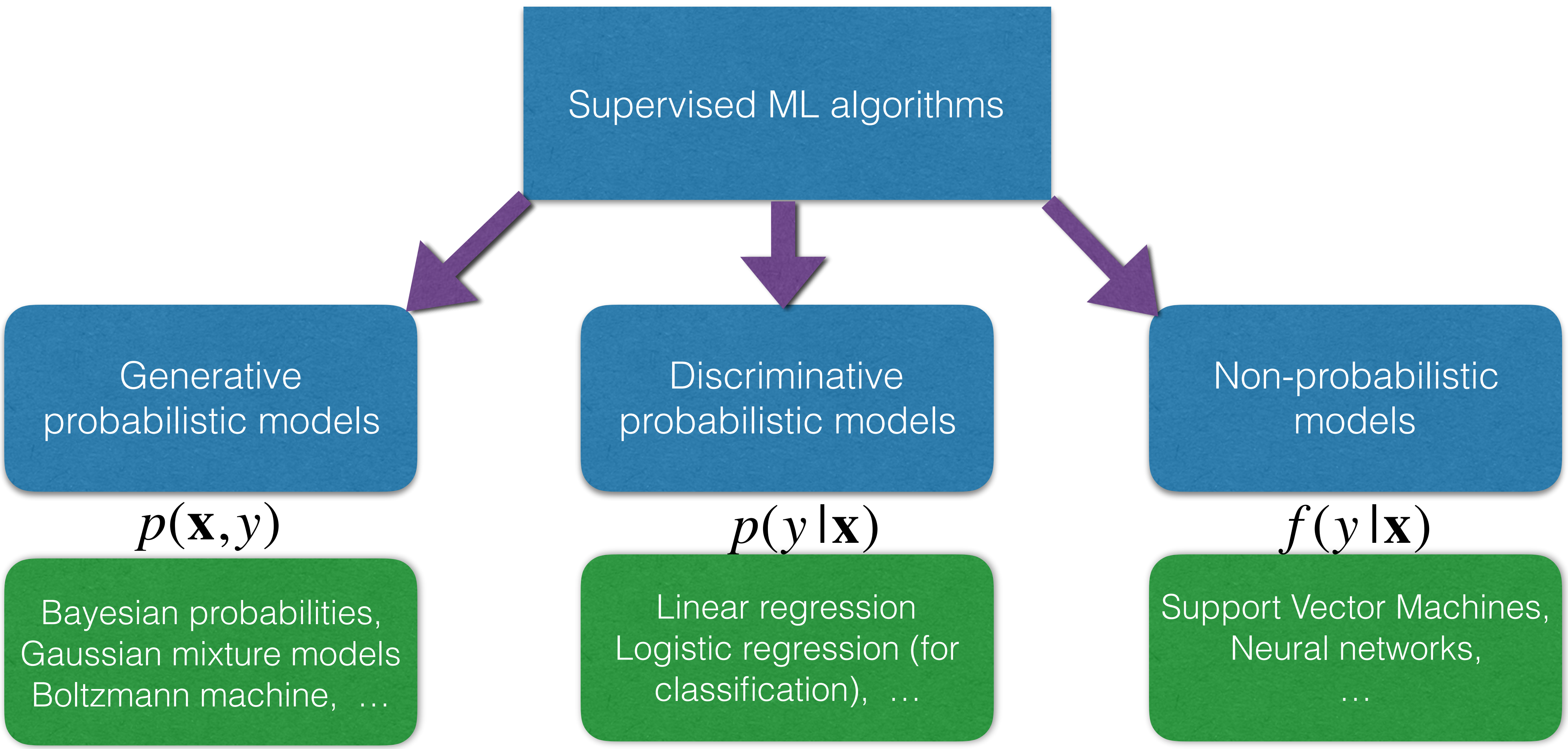
NYU Tandon School of Engineering, 2017

Supervised Learning algorithms

Most, but not all, supervised ML algorithms amount to **estimating a probability distribution** $p(y|\mathbf{x})$

Example: Linear Regression is equivalent to a **<u>discriminative probabilistic model</u>**

$$p(y|\mathbf{x}) = \mathcal{N}(y; \Theta^T \mathbf{x}; \sigma^2)$$



Can simulate from generative models

Probabilistic classification models

For probabilistic **regression** models, we estimate a probability distribution $p(y \mid \mathbf{x}), \quad y \in \mathbb{R}$

Probabilistic **classification** models:

Class labels

$$p(y \mid \mathbf{x}), \quad y \in \left(C_1, \ldots, C_K\right)$$

$\sum_k p(C_k \mid \mathbf{x}) = 1$

K classes

Probabilistic classification models

For probabilistic **regression** models, we estimate a probability distribution $p(y|\mathbf{x}), \quad y \in \mathbb{R}$

Probabilistic **classification** models:

Class labels

$$p(y|\mathbf{x}), \quad y \in \left(C_1, \ldots, C_K\right)$$

$\sum_k p(C_k|\mathbf{x}) = 1$

K classes

Binary classification (K=2):
Here $y = \{0,1\}$, or $y = \{-1,1\}$, or $y = \{low\,risk,\ high\,risk\}$, e.t.c.
Only one unknown: the “positive class” probability

$$p(y=1|\mathbf{x})$$

with $p(y=0|\mathbf{x}) = 1 - p(y=1|\mathbf{x})$

Probabilistic classification models

For probabilistic **regression** models, we estimate a probability distribution $p(y \mid \mathbf{x}), \quad y \in \mathbb{R}$

Probabilistic **classification** models:

Class labels

$$p(y \mid \mathbf{x}), \quad y \in \left(C_1, \ldots, C_K\right)$$

$\sum_k p(C_k \mid \mathbf{x}) = 1$

K classes

Binary classification (K=2):
Here $y = \{0,1\}$, or $y = \{-1,1\}$, or $y = \{low\,risk,\ high\,risk\}$, e.t.c.
Only one unknown: the “positive class” probability

$$p(y = 1 \mid \mathbf{x})$$

with $p(y = 0 \mid \mathbf{x}) = 1 - p(y = 1 \mid \mathbf{x})$

<u>Examples</u> of binary classification: bank failures, mortgage defaults, credit card fraud, anti-money laundering, price direction prediction

Logistic regression

Recall that a Linear Regression model fits a linear function

$$y = \boldsymbol{\Theta}^T \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^N, \boxed{y \in \mathbb{R}}$$

This is *not* on its own suitable to describe probabilities (as they should be numbers between 0 and 1).

Logistic regression

Recall that a Linear Regression model fits a linear function

$$y = \mathbf{\Theta}^T \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^N, \boxed{y \in \mathbb{R}}$$

This is *not* on its own suitable to describe probabilities (as they should be numbers between 0 and 1). **But!**

Let us apply a "squashing" function to y:

$$y \rightarrow h(y), \quad h: \mathbb{R} \rightarrow [0,1]$$

Nonlinear function

Logistic regression

Recall that a Linear Regression model fits a linear function

$$y = \mathbf{\Theta}^T \mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^N, \boxed{y \in \mathbb{R}}$$

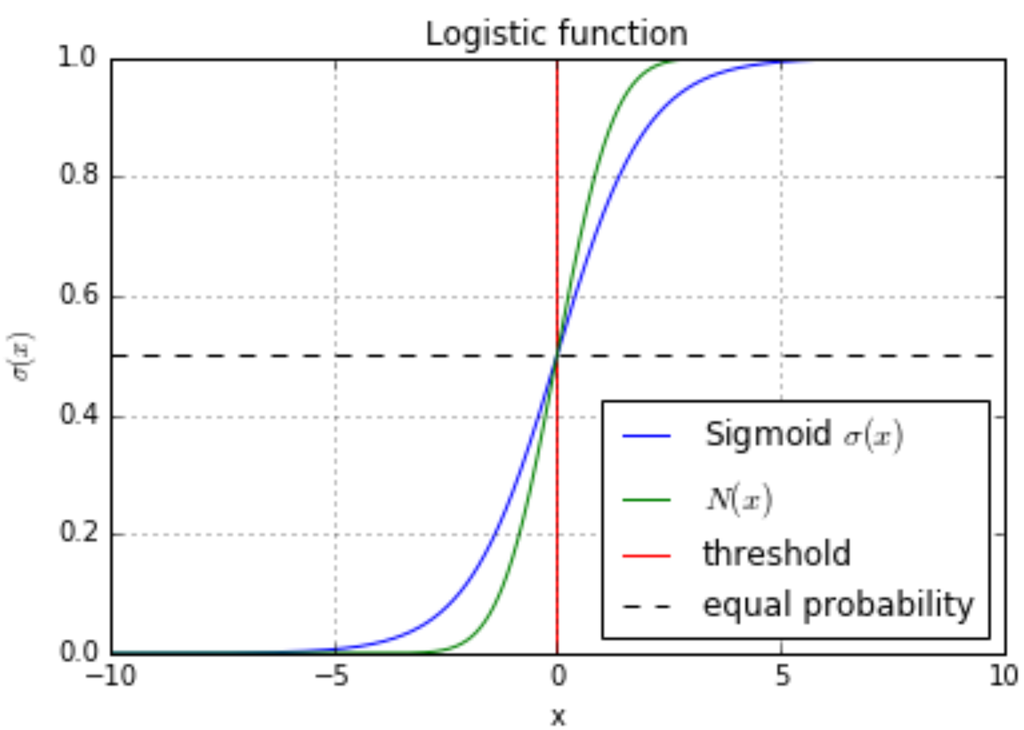
This is *not* on its own suitable to describe probabilities (as they should be numbers between 0 and 1). But!

Let us apply a “squashing” function to y :

Nonlinear function

$$y \to h(y), \quad h : \mathbb{R} \to [0,1]$$

Logistic Regression: take $h(y) = \sigma(y) = \dfrac{1}{1+\exp(-y)} \quad \Rightarrow$



$$\boxed{p(y=1 \mid \mathbf{x}) = \sigma(\theta^{\mathrm{T}}\mathbf{x}) = \frac{1}{1+\exp(-\theta^{\mathrm{T}}\mathbf{x})}}$$

Can also be derived from other models

MLE for Logistic Regression

Logistic Regression:

$$p_n(\theta) = p\left(y_n = 1 \mid \mathbf{x}_n\right) = \sigma(\theta^{\mathbf{T}}\mathbf{x}_n) = \frac{1}{1+\exp(-\theta^{\mathbf{T}}\mathbf{x}_n)}$$

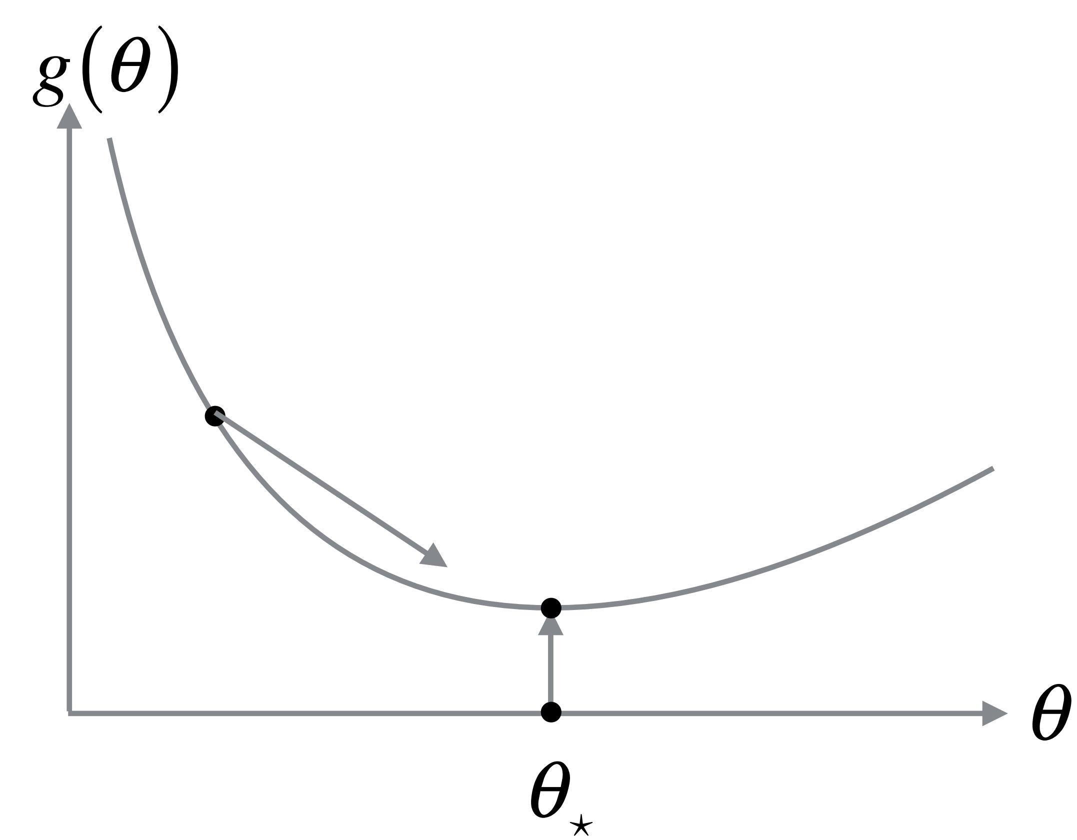
Likelihood:

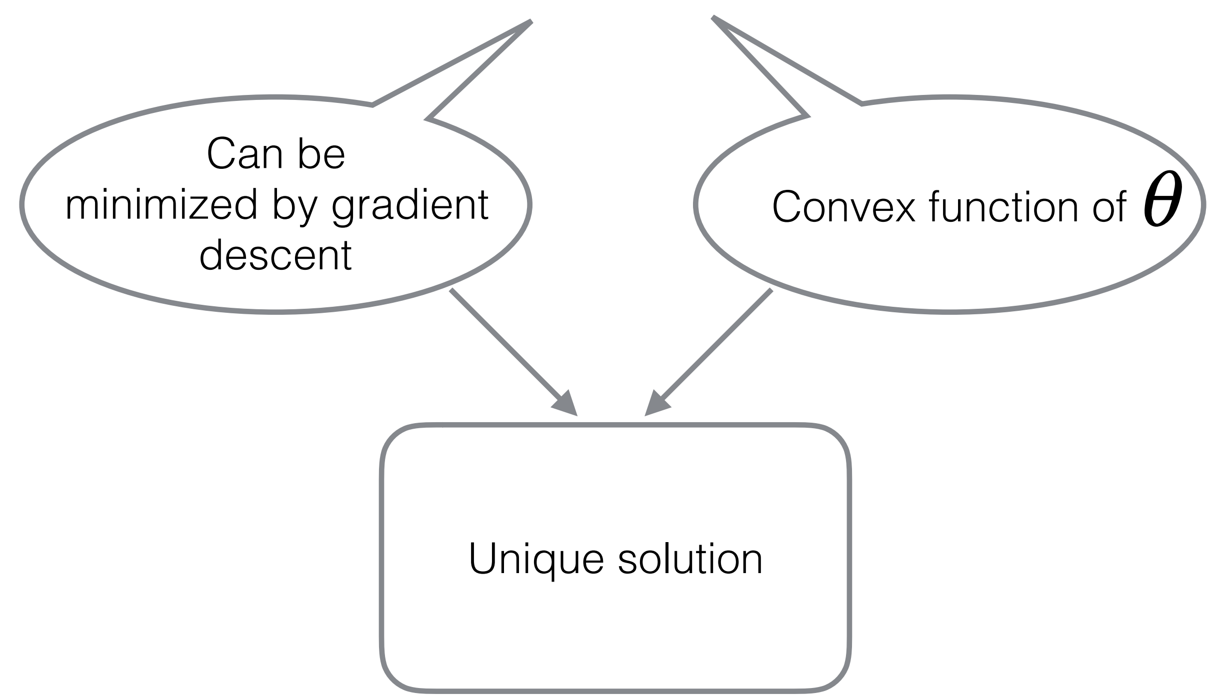
$$p(\mathbf{D} \mid \mathbf{M}, \Theta) = \prod_{n=1}^{N} p_n(\theta)^{y_n} \left(1 - p_n(\theta)\right)^{1-y_n}$$

observed values {0,1}

Negative LL:

$$-\log p(\mathbf{D} \mid \mathbf{M}, \Theta) = \sum_{n-1}^{N} \left[y_n \log p_n(\theta) + (1 - y_n) \log\left(1 - p_n(\theta)\right) \right]$$





Control question

Q: Select all correct statements:

1. Discriminative Probabilistic models enable simulating from a model.
2. Logistic Regression is a special kind of linear regression used for logistics-related tasks in supply chains and the military.
3. The fitted parameters of Logistic Regression critically depend on the initial guess: for different initial guesses, the results will always be different.
4. For a binary classification (K=2), Logistic Regression works as explained only for <u>categorical</u> labels (i.e. labels that cannot be convert to numbers that could be numerically compared with each other, for example “red” and “blue”). If labels are <u>ordinal</u> (i.e. they can be converted into numbers that could be compared with each other, for example high risk/low risk), one should use Linear Regression instead.

Correct answers: None.