EECS 395/495 Machine Learning

Aggelos K. Katsaggelos

Joseph Cummings Professor
Northwestern University
Department of EECS
Department of Linguistics
Argonne National Laboratory
NorthShore University Health System
Evanston, IL 60208
http://ivpl.eecs.northwestern.edu

Non-linear regression: Logistic regression

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

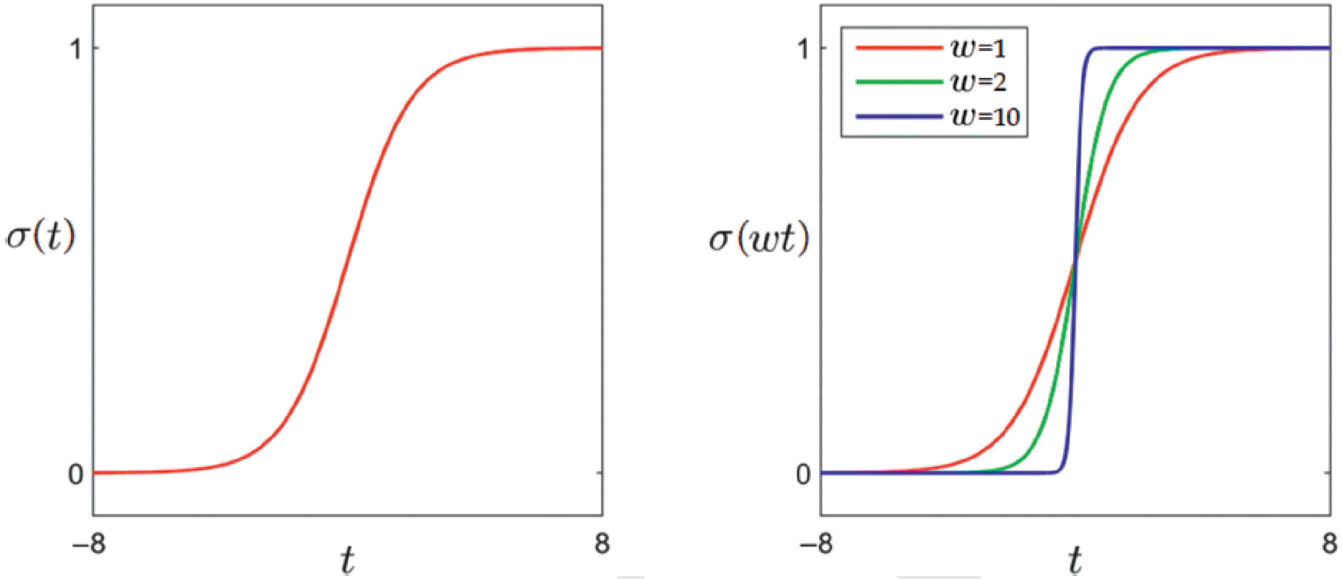


Fig. 3.10 (left panel) Plot of the logistic sigmoid function defined in (3.23). Note that the output of this function is always between 0 and 1. (right panel) By increasing the weight w of the sigmoid function $\sigma(wt)$ from $w = 1$ (red) to $w = 2$ (green) and finally to $w = 10$ (blue), the sigmoid becomes an increasingly good approximator of a "step function," that is a function that only takes on the values 0 and 1 with a sharp transition between the two.

Population growth model

f: current population level
(1-f): remaining capacity

$$\frac{df}{dt} = f(1 - f)$$

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

$$\sigma'(t) = \sigma(t)(1 - \sigma(t))$$

Non-linear regression: Logistic regression

Data distributed like a sigmoid, i.e., non-linear in x, w

$$\sigma\left(b + \mathbf{x}_p^T \mathbf{w}\right) \approx y_p, \quad p = 1, \ldots, P.$$

Non-convex LS function

$$g(b, \mathbf{w}) = \sum_{p=1}^{P} \left(\sigma\left(b + \mathbf{x}_p^T \mathbf{w}\right) - y_p\right)^2$$

$$\tilde{\mathbf{x}}_p = \begin{bmatrix} 1 \\ \mathbf{x}_p \end{bmatrix} \text{ and } \tilde{\mathbf{w}} = \begin{bmatrix} b \\ \mathbf{w} \end{bmatrix}.$$

$$\Rightarrow \quad \nabla g(\tilde{\mathbf{w}}) = 2\sum_{p=1}^{P} \left(\sigma\left(\tilde{\mathbf{x}}_p^T \tilde{\mathbf{w}}\right) - y_p\right) \sigma\left(\tilde{\mathbf{x}}_p^T \tilde{\mathbf{w}}\right) \left(1 - \sigma\left(\tilde{\mathbf{x}}_p^T \tilde{\mathbf{w}}\right)\right) \tilde{\mathbf{x}}_p.$$

Non-linear regression: Logistic regression

$$
\begin{aligned}
\tilde{\mathbf{w}}^{(k)} &= \tilde{\mathbf{w}}^{(k-1)} - \alpha_k \nabla g\left(\tilde{\mathbf{w}}^{(k-1)}\right) \\
&= \tilde{\mathbf{w}}^{(k-1)} - 2\alpha_k \sum_{p=1}^{P} \left(\sigma_p^{k-1} - b_p\right) \sigma_p^{k-1} \left(1 - \sigma_p^{k-1}\right) \tilde{\mathbf{x}}_p
\end{aligned}
$$

Bacterial Growth

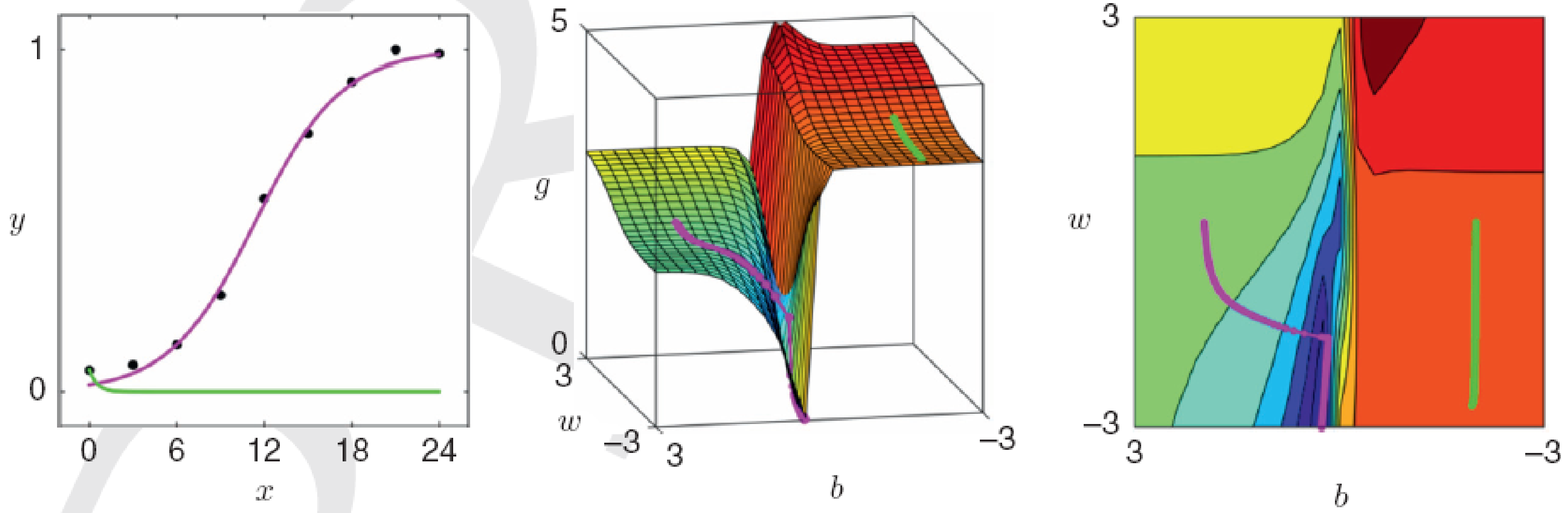
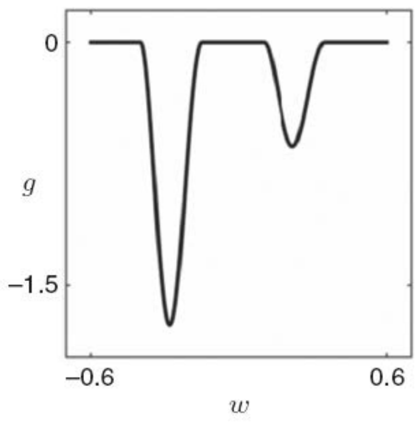


Fig. 3.11 (left panel) A dataset along with two sigmoidal fits (shown in magenta and green), each found via minimizing the Least Squares cost in (3.26) using gradient descent with a different initialization. A surface (middle) and contour (right) plot of this cost function, along with the paths taken by the two runs of gradient descent. Each path has been colored to match the resulting sigmoidal fit produced in the left panel (see text for further details). Data in this figure is taken from [48].

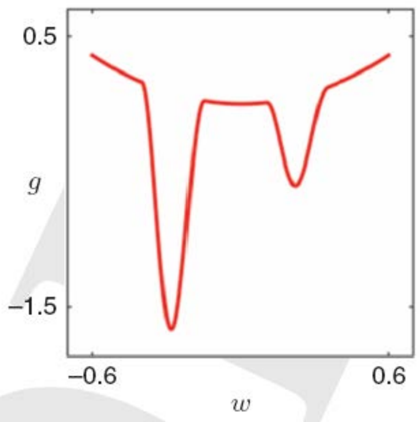
l_2 regularization

$$g\left(b, \mathbf{w}\right) + \lambda \left\|\mathbf{w}\right\|_2^2 .$$

$$g\left(w\right) = \max^2\left(0,\ \left(3w - 2.3\right)^3 + 1\right) + \max^2\left(0,\ \left(-3w + 0.7\right)^3 + 1\right).$$



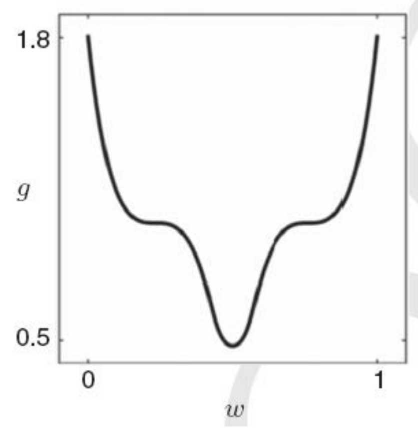
$\lambda = 1$



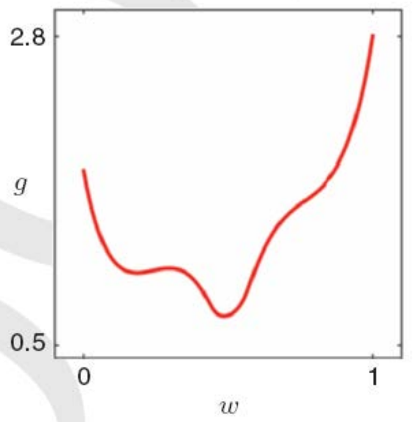
l_2 regularization

$$g\left(b, \mathbf{w}\right) + \lambda \left\| \mathbf{w} \right\|_2^2 .$$

$$g\left(w\right) = \max^2\left(0,\ \left(3w - 2.3\right)^3 + 1\right) + \max^2\left(0,\ \left(-3w + 0.7\right)^3 + 1\right).$$



$$\lambda = 1$$



l_2 regularization

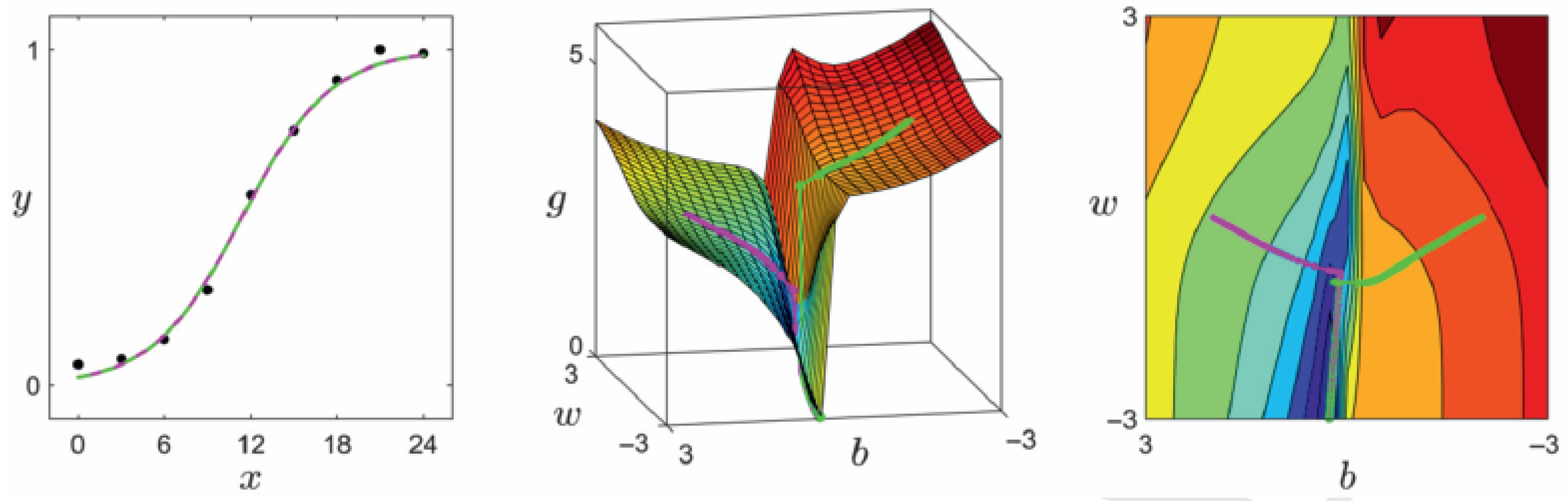


Fig. 3.13 A regularized version of Fig. 3.11. (left panel) Plot of the bacterial growth dataset along with two overlapping sigmoidal fits (shown in magenta and green) found via minimizing the ℓ_2 regularized Least Squares cost for logistic regression in (3.29) using gradient descent. (middle and right panels) The surface and contour plot of the regularized cost function along with the paths (in magenta and green) of gradient descent with same two initializations as shown in Fig. 3.11. While the surface is still non-convex, the large flat region that originally led the initialization of the green path to a poor solution with the unregularized cost has been curved upwards by the regularizer, allowing the green run of gradient descent to reach the global minimum of the problem. Data in this figure is taken from [48].

Regularized logistic regression

$$g(b, \mathbf{w}) = \sum_{p=1}^{P} \left(\sigma\left(b + \mathbf{x}_p^T \mathbf{w}\right) - y_p\right)^2 + \lambda \|\mathbf{w}\|_2^2$$

$$\tilde{\mathbf{x}}_p = \begin{bmatrix} 1 \\ \mathbf{x}_p \end{bmatrix} \text{ and } \tilde{\mathbf{w}} = \begin{bmatrix} b \\ \mathbf{w} \end{bmatrix} \Rightarrow g(\tilde{\mathbf{w}}) = \sum_{p=1}^{P} \left(\sigma\left(\tilde{\mathbf{x}}_p^T \tilde{\mathbf{w}}\right) - y_p\right)^2 + \lambda \|\mathbf{U}\tilde{\mathbf{w}}\|_2^2$$

with $\mathbf{U} = \begin{bmatrix} 0 & \mathbf{0}_{1\times N} \\ \mathbf{0}_{N\times 1} & \mathbf{I}_{N\times N} \end{bmatrix}$ $\qquad \mathbf{U}^T\mathbf{U} = \mathbf{U}$

$$\Rightarrow \nabla_{\tilde{\mathbf{w}}} g(\tilde{\mathbf{w}}) = 2\sum_{p=1}^{P} \left(\sigma\left(\tilde{\mathbf{x}}_p^T \tilde{\mathbf{w}}\right) - y_p\right) \sigma\left(\tilde{\mathbf{x}}_p^T \tilde{\mathbf{w}}\right) \left(1 - \sigma\left(\tilde{\mathbf{x}}_p^T \tilde{\mathbf{w}}\right)\right) \tilde{\mathbf{x}}_p + 2\lambda \mathbf{U}\tilde{\mathbf{w}}$$