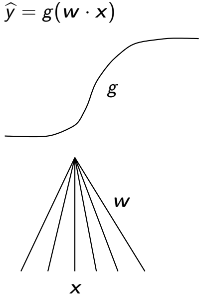
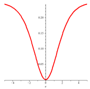
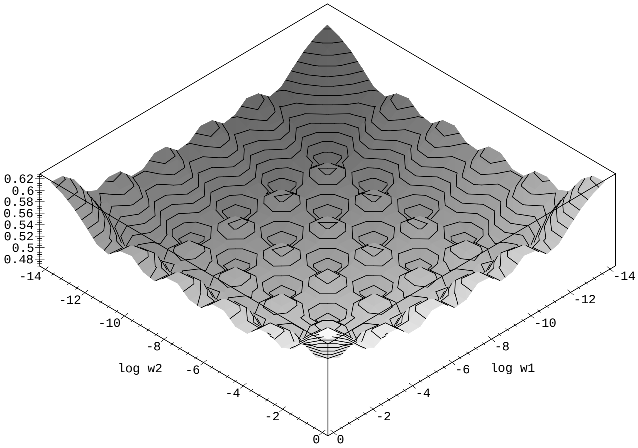
Learning with a single neuron

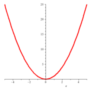


- Sigmoid function $g(z) = \frac{1}{1+e^{-z}}$
- For set of examples $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)$



total loss $\sum_{n=1}^{N} (g(\boldsymbol{w} \cdot \boldsymbol{x}_n) - y_n)^2/2$
can have exponentially # of minima in weight space [Bu,AHW]

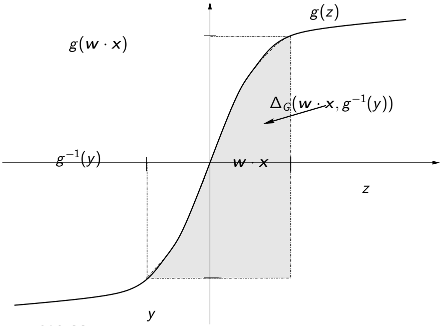




Want loss that is convex in ***w***

Convex plus convex is convex!

Bregman Divergences lead to good loss functions



transfer function $g = \nabla G$

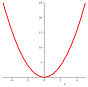
$$\begin{aligned}
\int_{g^{-1}(y)}^{\boldsymbol{w}\cdot\boldsymbol{x}} (g(z) - y)\, dz &= G(\boldsymbol{w}\cdot\boldsymbol{x}) - G(g^{-1}(y)) - (\boldsymbol{w}\cdot\boldsymbol{x} - g^{-1}(y))\, y \\
&= \Delta_G(\boldsymbol{w}\cdot\boldsymbol{x}, g^{-1}(y)) \\
\frac{\partial \Delta_G(\boldsymbol{w}\cdot\boldsymbol{x}, g^{-1}(y))}{\partial \boldsymbol{w}} &= (\underbrace{g(\boldsymbol{w}\cdot\boldsymbol{x})}_{\widehat{y}} - \underbrace{g(g^{-1}(y))}_{y})\boldsymbol{x}
\end{aligned}$$

Use $\Delta_G(\boldsymbol{w}\cdot\boldsymbol{x}, g^{-1}(y))$ as loss of $\boldsymbol{w}$ on $(\boldsymbol{x}, y)$

Called matching loss for g [AHW,HKW]

transfer func. $g(z)$	$G(z)$	match. loss $d_G(\boldsymbol{w}\cdot\boldsymbol{x}, g^{-1}(y))$
z	$\frac{1}{2}z^2$	$\frac{1}{2}(\boldsymbol{w}\cdot\boldsymbol{x} - y)^2$ square loss
$\frac{e^z}{1+e^z}$	$\ln(1+e^z)$	$y\ln y + (1-y)\ln(1-y)$ $+\ln(1+e^{\boldsymbol{w}\cdot\boldsymbol{x}}) - y\,\boldsymbol{w}\cdot\boldsymbol{x}$ logistic loss

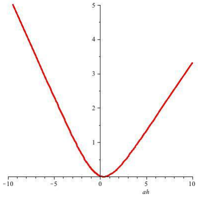
In all cases, $\frac{\partial \text{match.loss}}{\partial \boldsymbol{w}} = (\widehat{y} - y)\boldsymbol{x}$, where $\widehat{y} = g(\boldsymbol{w}\cdot\boldsymbol{x})$



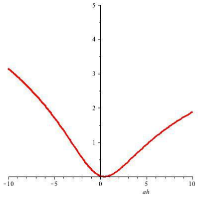
Matching loss always convex in $\boldsymbol{w}$

But convex losses NON-ROBUST TO OUTLIERS

Convex losses can't handle outliers

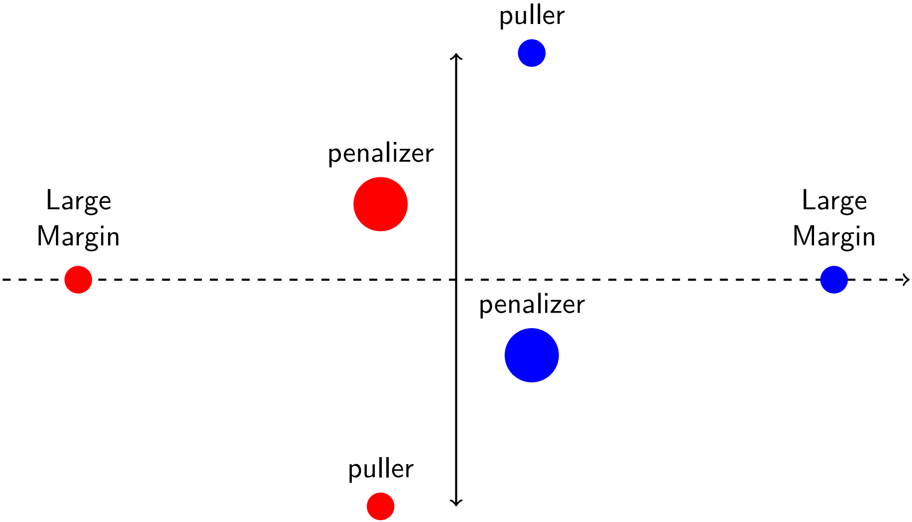


- Any convex loss grows at least linearly

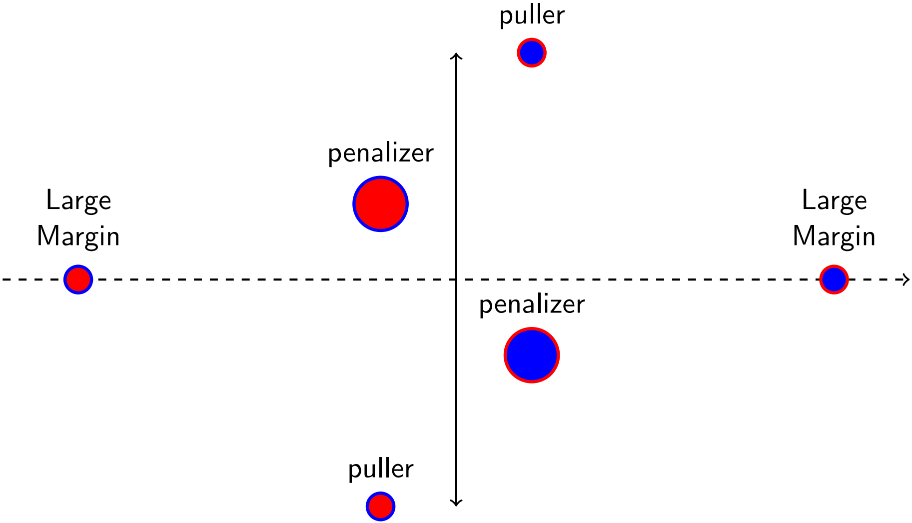


- We need the "wings to bend down", i.e. *forget / give up on* examples
- Non-convexity is needed to achieve robustness !!!

Fundamental example problem [LS,F]



Fundamental example problem [LS,F]



Fundamental example problem [LS,F]