Logistic Regression

The main tool of statistics is the Maximum-Likelihood method. Assuming the results of the experiment $\mathrm{x}^{(1)},\ \mathrm{x}^{(2)},\ \ldots,\ \mathrm{x}^{(N)}$, are samples from random variables $X^{(1)},\ X^{(2)},\ \ldots,\ X^{(N)}$ independent and identically distributed (i.i.d.) with pdf (or pmf) $p(x)$, the joint probability of the experiment is given by

$$L(\mathrm{x}^{(1)},\ \mathrm{x}^{(2)},\ \ldots,\ \mathrm{x}^{(N)}) = p(\mathrm{x}^{(1)})p(\mathrm{x}^{(2)})\ldots p(\mathrm{x}^{(N)}). \tag{1}$$

Problem 1. Consider the following random variable:

$Y\|X=x$	-1	1
P	$1-p$	p

Its pmf is given by $p(y) = [y=-1](1-p) + [y=1]p$. We use the sigmoid function to transform the distance to probability:

$$\mathrm{P}(Y=1|X=x) = p = \sigma(\langle w, \tilde{x}\rangle) = \frac{e^{w^T\tilde{x}}}{1+e^{w^T\tilde{x}}}.$$

If the true label is $y=1$, then

$$p = \frac{e^{w^T\tilde{x}}}{1+e^{w^T\tilde{x}}} = \frac{1}{1+e^{-w^T\tilde{x}}} = \frac{1}{1+e^{-yw^T\tilde{x}}}.$$

Prove that if $y=-1$, then $\mathrm{P}(Y=-1|X=x) = 1-p$ has the same expression.

Problem 2. Show that maximization of likelihood (1) leads to the loss function[1]

$$Loss = \frac{1}{N}\sum_{i=1}^{N} \log\left(1 + e^{-y^{(i)}\langle w,\ \tilde{x}^{(i)}\rangle}\right).$$

Hint: Maximization of L is equivalent to maximization of $\log L$ or to minimization of $-\log L$.

[1]This function is called *log-loss* or (negative) *binary cross-entropy loss.* Note that random variables $Y^{(i)}|X^{(i)} = x^{(i)}$ are not identically distributed (p depends on $x^{(i)}$), but we assume they are conditionally independent.