1 (e) This observation is reminiscent of the notion of support vectors in SVMs, where only a few training points contribute to the solution. In SVMs this is acheived using the hinge loss

$$\ell(y_i ; f(x_i)) = \max(0, 1 - y_i f(x_i))$$ instead of the logistic loss.

Because this loss has zero gradient whenever $y_i f(x_i) > 1$

the associated training pairs do not contribute at all to the optimization problem, which can be leveraged for efficient numerical optimization. It would thus be tempting to use:

$$r(y_i ; f(x_i)) = \exp(-\ell(y_i ; f(x_i)))$$ as a likelihood.

$$= \exp(-\max(0, 1 - y_i f(x_i)))$$

as a likelihood in our GP classification model. Unfortunately, this is not a valid likelihood.

To make this clear show that ; there is no constant c such that.

$$\sum_{y_i \in \{-1, 1\}} c \, r(y_i ; f(x_i)) = 1 \qquad \forall f(x_i) \in \mathbb{R}$$

thus $r(y_i ; f(x_i))$ is not a family of probability distributions, and can't be a likelihood.

Answer: We consider the normalization of suggested likelihood.

$$\sum_{y_i \in \{-1,1\}} c\, r(y_i ; f(x_i))$$

where $r(y_i ; f(x_i)) = \exp\left[-\max(0, 1 - y_i f(x_i))\right]$

we separate out +ve and -ve classifications.

s.t.

$$\sum_{\text{where } y_i=-1} c \exp\left[-\max(0, 1+f(x_i))\right] + \sum_{y_i=1} c \exp\left[-\max(0,1-f(x_i))\right]$$

Without loss of generality consider training points in +ve and negative classes with same $f(x_i) = f'$ value f'. Then either $f' \geq 1$ or $f' \leq 1$

Let $f' > 1$

without loss of generality consider training points in the +ve and -ve classes with $f(x_i) = f'$

then for two such points we have.

$$S = \underbrace{c \exp\left[-\max(0, 1+f')\right]}_{y_i=-1} + \underbrace{c \exp\left[-\max(0,1-f')\right]}_{y_i=1}$$

without loss of generality let $f' > 1$

then $\max(0, 1+f') = 1+f'$

$\max(0, 1-f') = 0$

$$S = c\left[\exp\left[-(1+f')\right] + e^{0}\right]$$

$$= c\left[1 + e^{-(1+f')}\right]$$

The only normalizer for S is when

$$c = \frac{1}{1+e^{-(1+f')}}$$

but this is not independent of $f(x_i)$

thus $r(y_i, f(x_i))$ is not a family of probability distributions, and can't be a likelihood.