1. **仅仅对输入抽取特征。**即特征函数为 $\mathbf{f}(\mathbf{x})$

2. 对输入和输出同时抽取特征。即特征函数为 $\mathbf{f}(\mathbf{x}, y)$

要看清二者的关系，一个简单的办法就是去考察题主提到的**最大熵模型和 logistic 回归模型。**确切地说，看看怎么把最大熵模型推导成 logistic 回归模型就可以了。

最大熵模型定义了在给定输入变量 $\mathbf{x}$ 时，输出变量 y 的条件分布：

$$P(y|\mathbf{x}, \boldsymbol{\theta}) = \frac{\exp\left(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y)\right)}{\sum_{\mathbf{y} \in Dom(y)} \exp\left(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y)\right)}$$

此处 $Dom(y)$ 是 y 所有可能取值的集合。

如果我们限定 y 为二元变量，即 $Dom(y) = \{y_0, y_1\}$，那么就可以把最大熵模型转换为 logistic 回归模型。我们还需要定义特征函数为

$$\mathbf{f}(\mathbf{x}, \mathbf{y}) = \begin{cases} \mathbf{g}(\mathbf{x}) & y = y_1 \\ \mathbf{0} & y = y_0 \end{cases}$$

即仅在 $y = y_1$ 时抽取 $\mathbf{x}$ 的特征。在 $y = y_0$ 时不抽任何特征（直接返回全为 0 的特征向量）。

将这个特征函数带回最大熵模型，我们得到当 $y = y_1$ 时

$$\begin{aligned}
P(y_1|\mathbf{x}) &= \frac{\exp\left(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y_1)\right)}{\exp\left(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y_0)\right) + \exp\left(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y_1)\right)} && \text{最大熵模型定義} \\
&= \frac{\exp\left(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right)}{\exp\left(\boldsymbol{\theta} \cdot \mathbf{0}\right) + \exp\left(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right)} && \text{特徵函數 } \mathbf{f} \text{ 的定義} \\
&= \frac{\exp\left(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right)}{1 + \exp\left(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right)} && e^0 = 1 \\
&= \frac{1}{\exp\left(-\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right) + 1} && \text{分子分母同除以} \exp\left(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right) \\
&= \sigma\left(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right) && \text{logistic 函數定義}
\end{aligned}$$

当 $y = y_0$ 时

$$
\begin{aligned}
P(y_0|\mathbf{x}) &= \frac{\exp\left(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y_0)\right)}{\exp\left(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y_0)\right) + \exp\left(\boldsymbol{\theta} \cdot \mathbf{f}(\mathbf{x}, y_1)\right)} \\
&= \frac{\exp\left(\boldsymbol{\theta} \cdot \mathbf{0}\right)}{\exp\left(\boldsymbol{\theta} \cdot \mathbf{0}\right) + \exp\left(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right)} \\
&= \frac{1}{1 + \exp\left(\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right)} \\
&= \frac{\exp\left(-\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right)}{\exp\left(-\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{x})\right) + 1} \\
&= 1 - P(y_1|\mathbf{x})
\end{aligned}
$$

我们发现，当类标签（class label）只有两个的时候，最大熵模型就是 logistic 回归模型。

表面上看，logistic 回归模型里面的特征函数的确只考虑 $\mathbf{x}$ 不考虑 y 。然而通过上面的推导，我们发现其实 $\mathbf{g}$ 抽取的特征仅仅在 $y = y_1$ 时被用到。

另外，logistic 回归模型当然有特征的概念。给模型一句自然语言，它肯定不认识。我们必须抽出像 *n* 元组（*n*-gram）、 词性（part-of-speech tag）等特征，才能把数据传给模型。特征函数无非就是模型的「眼睛」。