Linear model

$$\hat{y} = ax + b$$

→ When $x = 0$: $\hat{y} = b$

→ When $x \to x + 1$: $\hat{y} \to \hat{y} + a$

Normal linear model

$$\hat{y} = ax + b$$

$$y \sim Normal(\hat{y}, \sigma^2)$$

Likelihood

Definition:

$$\mathbf{L}(a, b \mid x, y) = P(y \mid a, b, x)$$

Poisson linear model

$$\log \hat{y} = ax + b$$

$$y \sim Poisson(\hat{y})$$

GLM: Generalized linear model

$$f(\hat{y}) = ax + b$$

$$y \sim \text{Distribution}(\hat{y}, \ldots)$$

Link functions f:

→ **Normal model**: $f(z) = z$

→ **Poisson model**: $f(z) = \log z$

Logistic model

or Binomial linear model

$$\log \frac{\hat{y}}{1-\hat{y}} = a_1 x_1 + a_2 x_2 + a_3 x_3$$

$$y \sim Binomial(1, \hat{y})$$

Logistic function

$$z = a_1 x_1 + a_2 x_2 + a_3 x_3$$

$$\log \frac{\hat{y}}{1 - \hat{y}} = z \Rightarrow$$

$$\hat{y} = \frac{1}{1 + e^{-z}}$$

Logistic model gradients

```
X # (n, 3)
Y # (n,)
coefs # (3, )
Z = (X * coefs).sum(axis=1) # (n,)
Yhat = 1/(1 + np.exp(-Z)) # (n,)
δ = Yhat - Y # (n,)
#   (3, n) @ (n,) -> (3,)
grad = X.T @ δ / nsamples # (3,)
```

Softmax model

or Multinomial linear model

$$\log \frac{\hat{y_k}}{1 - \hat{y_k}} = \sum_{j=1}^{m} a_{k,j} x_j$$

$$\vec{y} \sim Multinomial(1, \hat{\vec{y}})$$

$$\left[\vec{y} = (y_0, \ldots, y_{ncats}),\ \hat{\vec{y}} = (\hat{y}_0, \ldots, \hat{y}_{ncats})\right]$$

Softmax function

$$z = \sum_{j=1}^{m} a_{k,j} x_j$$

$$\log \frac{\hat{y}}{1 - \hat{y}} = z \Rightarrow$$

$$\hat{y}_k = \frac{e^{z_k}}{\sum_{i=1}^{ncats} e^{z_i}}$$

Softmax model gradients

```
X # (n, m )
Y # (n, ncats)
W # (m, ncats)
# (n, m ) @ (m, ncats) -> (n, ncats)
Z =     X @ W
Yhat = softmax(Z) # (n, ncats)
δ = Yhat - Y # (n, ncats)
# (m, n) @ (n, ncats) -> (m, ncats)
dW = X.T @ δ / nsamples # (m, ncats)
```

FFN model: two layers

$$Z_1 = X_1 W_1$$

$$X_2 = ReLU(Z_1)$$

$$Z_2 = X_2 W_2$$

$$\widehat{Y} = Softmax(Z_2)$$

$$Y \sim Multinomial(1, \widehat{Y})$$