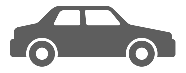
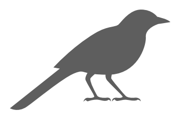
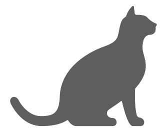
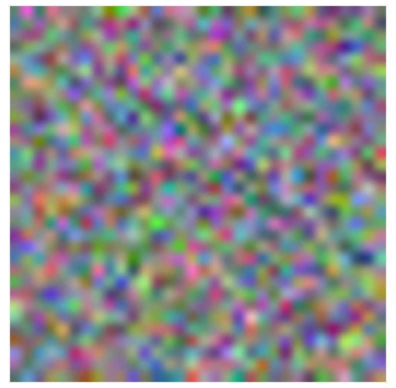
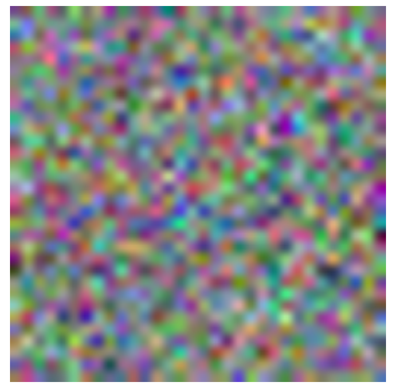
				
$W_{4\times32\times32\times3}$ $(\tilde{W}_{3072\times4})$ $X_{5\times32\times32\times3}$ $(\tilde{X}_{5\times3072})$				
	0.815	1.065	1.069	**-0.941**
	0.504	1.884	**1.517**	-1.221
	**0.763**	1.946	0.956	-0.930
	-0.165	**0.905**	0.383	0.358
	**0.748**	1.456	1.781	-0.905

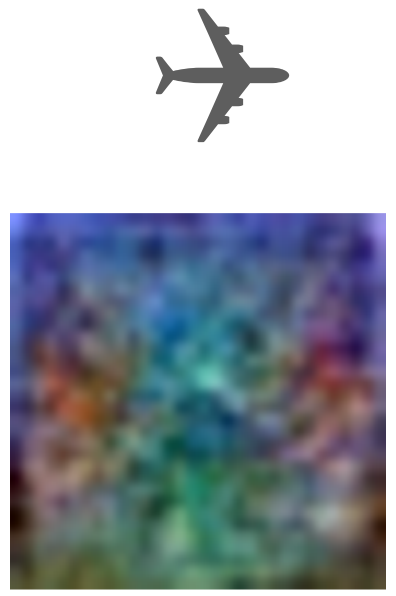
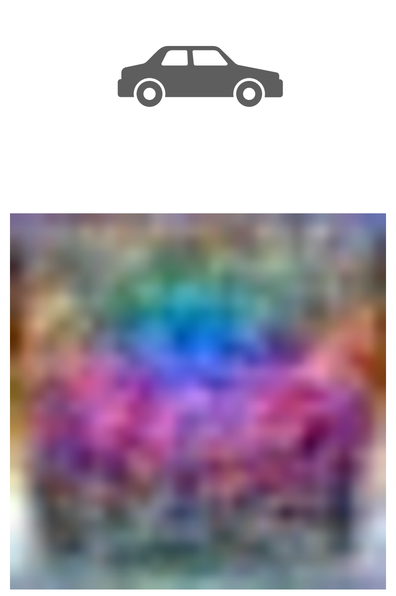
$$\begin{cases} f(X,W) = X \cdot W + b \\ L = \frac{1}{N}\sum_i L_i\left(f(x_i,W),y_i\right) \\ L_i = \sum_{j\neq y_i} max(0, s_j - s_{y_i} + 1) \end{cases}$$

$$\begin{cases} \frac{\partial L}{\partial W} = \frac{1}{N}\sum_i \left[\frac{\partial L_i\left(f(x_i,W),y_i\right)}{\partial f(x_i,W)} \cdot \frac{\partial f(x_i,W)}{\partial W}\right] \\ \frac{\partial L_i}{\partial W_j} = 1(s_j - s_{y_i} + 1 > 0)x_i \\ \frac{\partial L_i}{\partial W_{y_i}} = -\left(\sum_{j\neq y_i} 1(s_j - s_{y_i} + 1 > 0)\right)x_i \end{cases}$$

$$S = f(X,W) = \tilde{X}\cdot\tilde{W} + b = \begin{bmatrix} 0.815 & 1.065 & 1.069 & \mathbf{-0.941} \\ 0.504 & 1.884 & \mathbf{1.517} & -1.221 \\ \mathbf{0.763} & 1.946 & 0.965 & -0.930 \\ -0.165 & \mathbf{0.905} & 0.383 & 0.358 \\ \mathbf{0.748} & 1.456 & 1.781 & -0.905 \end{bmatrix} \quad y = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad S_y = \begin{bmatrix} -0.941 \\ 1.517 \\ 0.763 \\ 0.905 \\ 0.748 \end{bmatrix}$$

$$\max(0, S - S_y + 1) = \max\left(0, \begin{bmatrix} 2.756 & 3.006 & 3.010 & \mathbf{1} \\ -0.013 & 1.368 & \mathbf{1} & -1.737 \\ \mathbf{1} & 2.184 & 1.202 & -0.693 \\ -0.070 & \mathbf{1} & 0.477 & 0.452 \\ \mathbf{1} & 1.708 & 2.033 & -0.652 \end{bmatrix}\right) = \begin{bmatrix} 2.756 & 3.006 & 3.010 & \mathbf{1} \\ 0 & 1.368 & \mathbf{1} & 0 \\ \mathbf{1} & 2.184 & 1.202 & 0 \\ 0 & \mathbf{1} & 0.477 & 0.452 \\ \mathbf{1} & 1.708 & 2.033 & 0 \end{bmatrix}$$

$$L = \frac{1}{N}\left(\sum \max(0, S - S_y + 1) - N\right) = 3.639 \qquad \frac{\partial L}{\partial W} = \tilde{X}^T \cdot \left(\begin{bmatrix} 1 & 1 & 1 & \mathbf{1} \\ 0 & 1 & \mathbf{1} & 0 \\ \mathbf{1} & 1 & 1 & 0 \\ 0 & \mathbf{1} & 1 & 1 \\ \mathbf{1} & 1 & 1 & 0 \end{bmatrix} - \begin{bmatrix} & & & \mathbf{4} \\ & & \mathbf{2} & \\ \mathbf{3} & & & \\ & \mathbf{3} & & \\ \mathbf{3} & & & \end{bmatrix}\right)$$

$W_{4\times32\times32\times3}$ ($\tilde{W}_{3072\times4}$) \ $X_{5\times32\times32\times3}$ ($\tilde{X}_{5\times3072}$)				
	0.461	-1.016	0.516	**0.040**
	0.786	-1.502	**0.469**	0.258
	**1.227**	-0.985	0.498	-0.732
	0.282	**1.573**	-1.418	-0.433
	**0.896**	-0.927	-0.124	0.166

$$\begin{cases} f(X,W) = X\cdot W + b \\ L = \frac{1}{N}\sum_i L_i\left(f(x_i,W),y_i\right) \\ L_i = \sum_{j\neq y_i} max(0, s_j - s_{y_i} + 1) \end{cases}$$

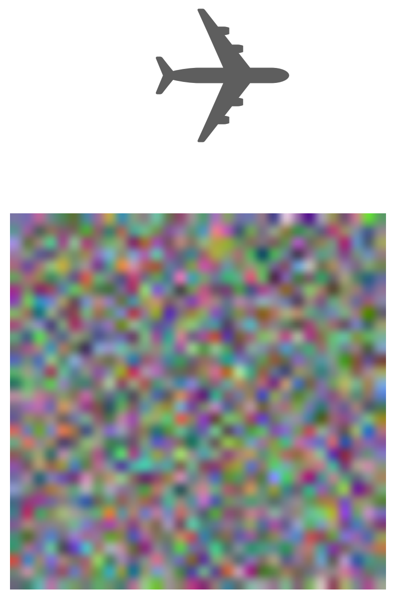
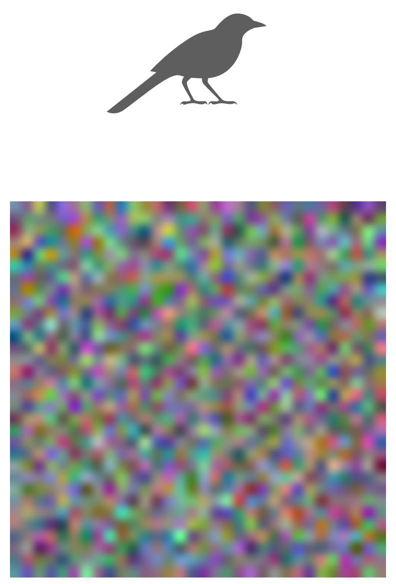
$$\begin{cases} \frac{\partial L}{\partial W} = \frac{1}{N}\sum_i \left[\frac{\partial L_i\left(f(x_i,W),y_i\right)}{\partial f(x_i,W)}\cdot\frac{\partial f(x_i,W)}{\partial W}\right] \\ \frac{\partial L_i}{\partial W_j} = 1(s_j - s_{y_i} + 1 > 0)x_i \\ \frac{\partial L_i}{\partial W_{y_i}} = -\left(\sum_{j\neq y_i} 1(s_j - s_{y_i} + 1 > 0)\right)x_i \end{cases}$$

$$S = f(X,W) = \tilde{X}\cdot\tilde{W} + b = \begin{bmatrix} 0.461 & -1.016 & 0.516 & \mathbf{0.040} \\ 0.786 & -1.502 & \mathbf{0.469} & 0.258 \\ \mathbf{1.227} & -0.985 & 0.498 & -0.732 \\ 0.282 & \mathbf{1.573} & -1.418 & -0.433 \\ \mathbf{0.896} & -0.927 & -0.124 & 0.166 \end{bmatrix} \quad y = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad S_y = \begin{bmatrix} 0.040 \\ 0.469 \\ 1.227 \\ 1.537 \\ 0.896 \end{bmatrix}$$

$$\max(0, S - S_y + 1) = \max\left(0, \begin{bmatrix} 1.421 & -0.057 & 1.475 & \mathbf{1} \\ 1.316 & -0.971 & \mathbf{1} & 0.789 \\ \mathbf{1} & -1.211 & 0.271 & -0.959 \\ -0.291 & \mathbf{1} & -1.992 & -1.006 \\ \mathbf{1} & -0.823 & -0.020 & 0.270 \end{bmatrix}\right) = \begin{bmatrix} 1.421 & 0 & 1.475 & \mathbf{1} \\ 1.316 & 0 & \mathbf{1} & 0.789 \\ \mathbf{1} & 0 & 0.271 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0.270 \end{bmatrix}$$

$$L = \frac{1}{N}\left(\sum \max(0, S - S_y + 1) - N\right) = 1.109$$

$$\frac{\partial L}{\partial W} = \tilde{X}^T\cdot\left(\begin{bmatrix} 1 & 0 & 1 & \mathbf{1} \\ 1 & 0 & \mathbf{1} & 1 \\ \mathbf{1} & 0 & 1 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ \mathbf{1} & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} & & & \mathbf{3} \\ & & \mathbf{3} & \\ \mathbf{2} & & & \\ & \mathbf{1} & & \\ \mathbf{2} & & & \end{bmatrix}\right)$$

$W_{4\times32\times32\times3}$ $(\tilde{W}_{3072\times4})$ / $X_{5\times32\times32\times3}$ $(\tilde{X}_{5\times3072})$				
	0.267	0.343	0.344	**0.046**
	0.126	0.503	**0.348**	0.023
	**0.176**	0.576	0.216	0.032
	0.136	**0.398**	0.236	0.230
	**0.166**	0.337	0.466	0.032

$$\begin{cases} S = f(X, W) = X \cdot W + b \\ P(Y = y_i \mid X = x_i) = \frac{e^{s_k}}{\sum_j e^{s_j}} \\ L_i = -\,logP(Y = y_i \mid X = x_i) \end{cases}$$

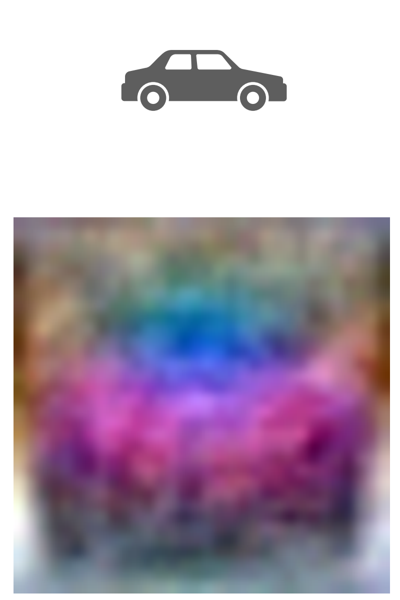
$$\begin{cases} \dfrac{\partial L_i}{\partial W_j} = -\dfrac{\sum_j e^{s_j}}{e^{s_{y_i}}} \cdot \dfrac{0 \cdot \sum_j e^{s_j} - e^{s_{y_i}} e^{s_j}}{(\sum_j e^{s_j})^2} \cdot x_i \\ \qquad = \dfrac{e^{s_j}}{\sum_j e^{s_j}} \cdot x_i \\ \qquad = P(Y = y_i \mid X = x_i) \cdot x_i \\ \dfrac{\partial L_i}{\partial W_{y_i}} = -\dfrac{\sum_j e^{s_j}}{e^{s_{y_i}}} \cdot \dfrac{e^{s_{y_i}} \sum_j e^{s_j} - e^{s_{y_i}} e^{s_{y_i}}}{(\sum_j e^{s_j})^2} \cdot x_i \\ \qquad = \dfrac{e^{s_{y_i}} - \sum_j e^{s_j}}{\sum_j e^{s_j}} \cdot x_i \\ \qquad = \big(P(Y = y_i \mid X = x_i) - 1\big) \cdot x_i \end{cases}$$

$$S = f(X, W) = \tilde{X} \cdot \tilde{W} + b = \begin{bmatrix} 0.815 & 1.065 & 1.069 & \mathbf{-0.941} \\ 0.504 & 1.884 & \mathbf{1.517} & -1.221 \\ \mathbf{0.763} & 1.946 & 0.965 & -0.930 \\ -0.165 & \mathbf{0.905} & 0.383 & 0.358 \\ \mathbf{0.748} & 1.456 & 1.781 & -0.905 \end{bmatrix} \qquad y = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$P(Y = y_i \mid X = x_i) = \frac{e^{s_{y_i}}}{\sum_j e^{s_j}} = \begin{bmatrix} 0.267 & 0.343 & 0.344 & \mathbf{0.046} \\ 0.126 & 0.503 & \mathbf{0.348} & 0.023 \\ \mathbf{0.176} & 0.576 & 0.216 & 0.032 \\ 0.136 & \mathbf{0.398} & 0.236 & 0.230 \\ \mathbf{0.166} & 0.337 & 0.466 & 0.032 \end{bmatrix}$$

$$L = \frac{1}{N}\left(\sum -\,logP(Y = y_i \mid X = x_i)\right) = 1.717$$

$$\frac{\partial L}{\partial W} = \tilde{X}^T \cdot \left(\begin{bmatrix} 0.267 & 0.343 & 0.344 & \mathbf{0.046} \\ 0.126 & 0.503 & \mathbf{0.348} & 0.023 \\ \mathbf{0.176} & 0.576 & 0.216 & 0.032 \\ 0.136 & \mathbf{0.398} & 0.236 & 0.230 \\ \mathbf{0.166} & 0.337 & 0.466 & 0.032 \end{bmatrix} - \begin{bmatrix} & & & \mathbf{1} \\ & & \mathbf{1} & \\ \mathbf{1} & & & \\ & \mathbf{1} & & \\ \mathbf{1} & & & \end{bmatrix} \right)$$

$W_{4\times32\times32\times3}$ $(\tilde{W}_{3072\times4})$ / $X_{5\times32\times32\times3}$ $(\tilde{X}_{5\times3072})$				
	0.345	0.068	0.392	**0.195**
	0.533	0.027	**0.259**	0.180
	**0.716**	0.038	0.196	0.050
	0.155	**0.767**	0.019	0.058
	**0.541**	0.058	0.186	0.215

$$\begin{cases} S = f(X, W) = X \cdot W + b \\ P(Y = y_i | X = x_i) = \frac{e^{s_k}}{\sum_j e^{s_j}} \\ L_i = -\,logP(Y = y_i | X = x_i) \end{cases}$$

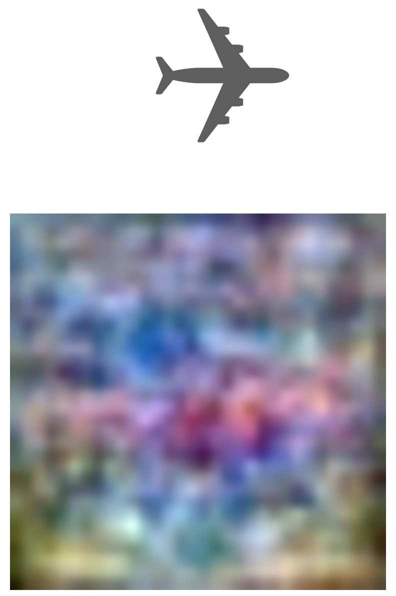
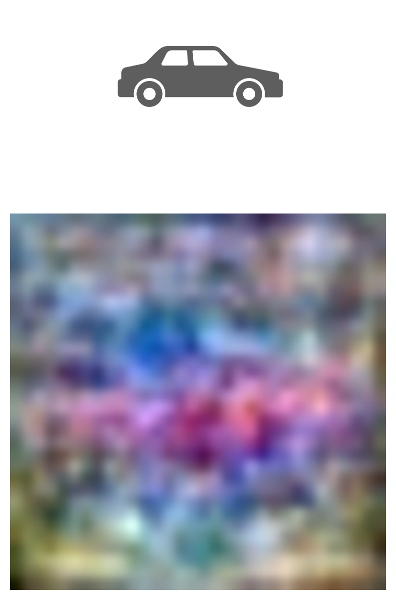
$$\begin{cases} \frac{\partial L_i}{\partial W_j} = -\frac{\sum_j e^{s_j}}{e^{s_{y_i}}} \cdot \frac{0 \cdot \sum_j e^{s_j} - e^{s_{y_i}} e^{s_j}}{(\sum_j e^{s_j})^2} \cdot x_i \\ \quad = \frac{e^{s_j}}{\sum_j e^{s_j}} \cdot x_i \\ \quad = P(Y = y_i | X = x_i) \cdot x_i \\ \frac{\partial L_i}{\partial W_{y_i}} = -\frac{\sum_j e^{s_j}}{e^{s_{y_i}}} \cdot \frac{e^{s_{y_i}} \sum_j e^{s_j} - e^{s_{y_i}} e^{s_{y_i}}}{(\sum_j e^{s_j})^2} \cdot x_i \\ \quad = \frac{e^{s_{y_i}} - \sum_j e^{s_j}}{\sum_j e^{s_j}} \cdot x_i \\ \quad = \left(P(Y = y_i | X = x_i) - 1\right) \cdot x_i \end{cases}$$

$$S = f(X, W) = \tilde{X} \cdot \tilde{W} + b = \begin{bmatrix} 0.517 & -1.105 & 0.644 & \mathbf{-0.052} \\ 1.203 & -1.767 & \mathbf{-0.483} & 0.119 \\ \mathbf{-1.727} & -1.204 & 0.431 & -0.929 \\ 0.372 & \mathbf{-1.969} & -1.709 & -0.618 \\ \mathbf{1.065} & -1.171 & -0.002 & 0.140 \end{bmatrix} \qquad y = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$P(Y = y_i | X = x_i) = \frac{e^{s_{y_i}}}{\sum_j e^{s_j}} = \begin{bmatrix} 0.345 & 0.068 & 0.392 & \mathbf{0.195} \\ 0.533 & 0.027 & \mathbf{0.259} & 0.180 \\ \mathbf{0.716} & 0.038 & 0.196 & 0.050 \\ 0.155 & \mathbf{0.767} & 0.019 & 0.058 \\ \mathbf{0.541} & 0.058 & 0.186 & 0.215 \end{bmatrix}$$

$$L = \frac{1}{N}\left(\sum -\,logP(Y = y_i | X = x_i)\right) = 0.839$$

$$\frac{\partial L}{\partial W} = \tilde{X}^T \cdot \left(\begin{bmatrix} 0.345 & 0.068 & 0.392 & \mathbf{0.195} \\ 0.533 & 0.027 & \mathbf{0.259} & 0.180 \\ \mathbf{0.716} & 0.038 & 0.196 & 0.050 \\ 0.155 & \mathbf{0.767} & 0.019 & 0.058 \\ \mathbf{0.541} & 0.058 & 0.186 & 0.215 \end{bmatrix} - \begin{bmatrix} & & & \mathbf{1} \\ & & \mathbf{1} & \\ \mathbf{1} & & & \\ & \mathbf{1} & & \\ \mathbf{1} & & & \end{bmatrix} \right)$$

$W_{4\times32\times32\times3}$ ($\tilde{W}_{3072\times4}$) / $X_{5\times32\times32\times3}$ ($\tilde{X}_{5\times3072}$)				
	0.408	0.011	0.547	**0.033**
	0.869	0.004	**0.122**	0.005
	**0.878**	0.002	0.117	0.003
	0.128	**0.848**	0.011	0.013
	**0.892**	0.012	0.089	0.008