	Finding the golps in MUDL
	min 0-1 loss < discrete function inteact optimisation
	inteact oftheir ation
	D-1 Loot even integui) Optimization problem
	At me Zoto Groldon
	Very difficult to solve
	So they want the surrect obtinisation
Γ	So they want the surrogete optimisation problems!
E	(f(x,0),y)
<u> </u>	/^ X , / \
	(9,y)~P(data) (xxxxx)
	* Lets assume that the dataset, and testing
	* Lets assume that the dataset, and testing one coming from the same probability afterwhetion
+	
	> We one never minimizing Enfectation before
	Caron VI (Scarp) 12 dad
	$\iint (p(x,y)) L(\mathcal{F}(x,b),y) dxdy$
	We are never dointhis (Were never talking
	We one never doint this (Were never talking before about gradient descert)

Instead, une one doing
$\frac{1}{2} \sum_{i=1}^{n} L(f(x_{i}; \theta), y)$
1) J=1 (not a kL clivergence
min $\Lambda = E(L(f(x_i, \theta), y))$ no joint chishibution)
(2,4)~ Polata
$\frac{\partial \mathcal{E}}{\partial \mathbf{Q}_{k}} = \frac{1}{h} \sum_{i=1}^{n} \frac{\partial \mathcal{E}_{i}}{\partial \mathbf{Q}_{k}} \qquad \text{ Jointuch}$
$\int_{\mathbb{R}^{n}} \int_{\mathbb{R}^{n}} \int_{$
Somuch
3 SCrD/Minibatch
The entimate de ly deron
OOK OBE
<u> </u>
Estimate 2 (E2 +65 + + (75 Eroso) minipatch gradient disent
Estimate 5 (42765 - 175 - 186)
minipatch judicit
botch gradient descent
botch gradient descent typically means classical
c'è bigge batch size, better estimate? (19 yes)
Minibatch stochastic, Online
most Frecent
Stuff (somples)
Stuff (somples) in timestamp or 5th for example
or 5th for example